A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Emotional Attitudes of Former Soviet Citizens as Studied by the Technique of Projective Questions 143
Helen Beier and Eugenia Hanfmann

Volunteering as an Avoidance Act 154
Robert R. Blake, Howard Berkowitz, Roy Q. Bellamy, and Jane Srygley Mouton

The Structure of Social Attitudes in Lebanon 157
E. Terry Prothro and Jack D. Keehn

Rorschach and TAT Indices of Homosexuality in Overt Homosexuals, Neurotics, and Normal Males 161
Anthony Davids, Mark Joelson, and Charles McArthur

The Effect of Psychopathology on Visual Discrimination 173
Herman Niebuhr, Jr., and David Cohen

The Rotation of Drawings by Brain-Damaged Patients 178
Aubrey J. Yates

Hope of Success and Fear of Failure as Aspects of Need for Achievement 182
Russell A. Clark, Richard Teevan, and Henry N. Ricciuti

Continued on inside front cover

Volume 53 SEPTEMBER 1956 Number 2

CONTENTS—*Continued*

Sexual Symbolism and the Variables of Sex and Personality Integration.... 187
Austin Jones

The Effects of Three Variables on Children's Concepts of Physical Causality. 191
Martin L. Nass

Speed of Form Recognition as a Function of Stimulus Factors and Test Anxiety..... 197
Frederick C. Rockett

Avoidance Conditioning of Verbal Behavior Without Awareness: A Paradigm of Repression..... 203
Charles W. Eriksen and James L. Kuethe

The Effects of Varying Combinations of Authoritarian and Equalitarian Leaders and Followers..... 210
William Haythorn, Arthur Couch, Don Haefner, Peter Langham, and Launor Carter

Paradoxical Effects in the Reaction of Schizophrenics..... 220
P. H. Venables and J. Tizard

Reaction Time and Response Amplitude as a Function of Anxiety and Stimulus Intensity..... 225
Alfred Castaneda

A Genetic Approach to the Interpretation and Evaluation of the Process-Reactive Distinction in Schizophrenia..... 229
Wesley C. Becker

Assessment of Abstract Behavior in a Non-Western Culture..... 237
Gustav Jahoda

The Manipulation of Dominance in Monkeys with Conditioned Fear..... 244
J. V. Murphy and R. E. Miller

CRITIQUE AND NOTES

An Investigation of Phonetic Symbolism..... 249
Irving Maltzman, Lloyd Morrisett, Jr., and Lloyd O. Brooks

Psychotherapy and the Recovery from Neurosis..... 251
Dalbir Bindra

A Short Form of Witkin's Embedded Figures Test..... 254
Douglas N. Jackson

Achievement Motivation and Learning..... 255
Patricia A. Karolchuch and Leonard Worell

The Child's Perception of the Parent..... 257
Jerome Kagan

A Further Study of the McKinney Reporting Test Scores..... 258
John R. Barry

Volume 53 SEPTEMBER 1956 Number 2

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, New York University

Associate Editor: E. J. SHOBEN, JR., Teachers College, Columbia University

Managing Editor: LORRAINE BOUTHILET

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study in the Behavioral Sciences

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D. C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1956 by the American Psychological Association, Inc

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis; those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area, this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscripti on price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

EMOTIONAL ATTITUDES OF FORMER SOVIET CITIZENS, AS STUDIED BY THE TECHNIQUE OF PROJECTIVE QUESTIONS[1]

HELEN BEIER
Boston University
AND EUGENIA HANFMANN
Brandeis University

IN STUDYING the personality patterns of former Soviet citizens we used, among other tests, the projective questions that had been employed in the California studies of authoritarianism (1). These questions were included mainly because, by inquiring about sentiments and cathexes, they elicit expressions of significant attitudes and provide rich material for personality study. Furthermore, since this test had been calibrated as a measure of authoritarianism, we planned to compare the Russian and the American performance in terms of frequency of responses indicating the authoritarian and the equalitarian personality patterns. Because of this double utilization of the data, all test responses were subjected to two separate procedures: they were scored as high, medium, or low in authoritarianism,[2] and, in an independent operation, they were classified according to their content. The selection of categories for this latter ordering was determined partly by the material itself, partly by our attempting to group the responses in ways that would permit a variety of psychological analyses. The content categories were different for each question and were not intended to reproduce the variables by which the California scoring is guided.

In the present communication we shall report only the results of the content analysis of the Russian and American responses, with the purpose of contributing to the description of the emotional and evaluative attitudes of the Russian subjects. The discussion of the scores of authoritarianism as such must be reserved for a later publication: it could not be presented without a detailed description of the various alternate ways in which we tried to meet the difficulties involved in applying a scoring system cross-culturally. It should be mentioned, however, that none of the variations of the scoring procedure we used yielded any large differences in the average scores of authoritarianism of the Russian and American samples. In the light of this finding, the differences in the actual content of the responses of the two groups emphasize the fact that both "high" and "low" scores can be given to responses indicating very different kinds of experience and behavior. Thus our results have some bearing on the general problem of subpatterns of authoritarianism and equalitarianism, a topic which will be taken up in the summarizing discussion of the findings.

[1] The authors are grateful to the Russian Research Center of Harvard University which sponsored this study under a grant from the Human Resources Research Institute of the U. S. Air Force. Drs. A. Inkeles and D. J. Levinson read the manuscript and made many helpful suggestions.

[2] The scoring in terms of authoritarianism was done by J. Orton and H. Beier with the advice and assistance of D. J. Levinson and Miss L. Heims.

METHOD

The Russian sample included 39 men and 9 women, almost all of them Great Russians. They were selected for a clinical study more or less at random from a much larger number of former Soviet citizens who were interviewed in Munich in the winter of 1950–51 by the Harvard Project on the Soviet Social System. Of this group of 48 subjects, 30 were brought to Germany during the war: either as war prisoners, or as civilians imported by the Germans to increase their labor force. Their reason for not returning to their native country after the war was fear of the suspicious and punitive attitude of the regime toward those who had spent a long period of time outside the Soviet Union. The remaining 18 subjects (all of them men) found their way to Germany after the war, most of them escaping from the Russian Army of Occupation; some of them were fairly recent arrivals, still very unsettled in their new environment. Younger people, who had grown up under the new regime, predominated in our group: 19 subjects were in their twenties, 19 in their thirties and only 10 older than forty. The middle occupational levels (white collar workers, technicians) predominated, with 23 subjects; the other half of the group were evenly divided between the professional or managerial level, 12, and that of skilled and unskilled laborers, including a few peasants, 13. The distribution of educational levels followed the same pattern of concentra-

tion in the middle (high school) level, 22, but the college graduates predominated over those who had had only 1–4 years of schooling, 16 and 10 subjects, respectively. About half of the subjects had been affiliated in the Soviet Union either with the Party or with the Communist Youth Organization.

The American group was matched to the Russian group, subject by subject, in terms of age, sex, educational level, and occupational grouping. All subjects came from homes where English was spoken, although, in a few cases, not as the only language; about half of them were born in Massachusetts, the rest being divided between other eastern states, the South and the Middle West. Naturally the two groups could not be matched in terms of their life experiences, recent and past. In both groups the participation in the study was voluntary, and the subjects were mostly recruited through organizations to which they belonged, and were paid for their time. However, for the Russians, the chance to tell the Americans about their lives and their grievances made participation a much more vital experience than it was for the American subjects.

The subjects were given the eight projective questions used in the California studies, and one additional question. (Because of some accidental omissions the number of Russian subjects varies for different questions between 45 and 48.) The questions were asked in the subjects' native language, with probing used when necessary, and with no time pressure. The answers were recorded by the tester. The test was preceded by interviews—very extensive ones in the case of the Russians—so that rapport had been well established. Most of the subjects responded to the questions naturally and willingly.

The content categories for each question were drawn up by the senior author on the basis of a preliminary survey of responses. Since these categories were purely factual, their application demanded no interpretation of responses and presented no particular difficulties; coding as done by the two authors produced very consistent results. If the discussion of a question by a subject contained several separate points they could be entered under different content categories; the term "response" as used by us applies to these units of content. The Russians were more voluble than the Americans and gave a higher total number of responses in 7 out of 9 questions.

The statistical comparison of the two groups was done in terms of number of *subjects* giving responses in a given content category; the significance of the differences between the groups was tested by means of Zubin's nomographs (3).[3] The results of the content analysis will be presented mainly in terms of these comparisons, but some reference will be made also to the distribution of the total *responses* of each group over the various content categories. Whenever the content categories bear some resemblance to the specific qualitative variables that were mapped out in the California studies, these relationships will be indicated.

[3] We wish to thank Dr. T. Alper for her advice and help with the statistical work.

RESULTS

Question 1. All of us are sometimes in a bad mood. What feelings do you think are most unpleasant for you; which of them irritate (upset) you most of all?[4]

Of all nine questions this one elicited from the Russians the largest number of responses referring to vital deprivations and to insecurities inherent in their past and present situation: such references were present in two thirds of the total Russian responses. This high incidence is probably due both to the nature of the question and to its position at the beginning of the series: our subjects utilized this first question for pouring out their most pressing concerns. Responses in one or in both of the closely related categories: "frustration of vital needs" and "negative external conditions" were made by 28 Russians and by 8 Americans.[5] The category "no prospects in the future" was used by the Russians only (9 subjects**).

Some of the most frequent Russian responses given in these categories refer to lack of work and of vital necessities, to worry about the fate of the relatives who stayed in the USSR, and about one's own uncertain future. This last worry was increased by the fear of the Soviets' invading Germany which was widespread in Europe at the time of the study: such an event would spell disaster to the nonreturnees. Following are some illustrative quotations from the records:

> I am most disturbed when there is nothing to eat, no money to buy clothing, particularly if one of the children is sick. When all are in good health you can bear these lacks more easily. Right now I ought to buy shoes for my boy—he has not been able to go out of the house for a week. But if we buy clothing we need, not enough would be left for food.
>
> All the moral and physical sufferings that result from our abnormal life: absence of regular work which would provide personal satisfaction and the material basis for living. This problem is with us always: this futile chase after work.
>
> At home it was fear of getting involved with the

[4] In most of the questions we simplified the original formulations to make them more understandable to the subjects of lower educational level in both national groups; consequently the version given here is somewhat different from that used in the California studies.

[5] This difference is significant at the .01 level of probability. Such differences will be indicated in the text by double asterisks, those significant at the .05 level by single asterisks.

NKVD—how to escape them. Now I am disturbed by the feeling that war might start any day, that I might wake up and find the Red Army here; and this is a fear not just for myself.

The American discussions contain no comparable responses. Those few answers that have been categorized under "negative external conditions" refer not to vital but to trivial or even freakish situations: "I am upset when people smoke in my car;" "When somebody pours cold water on my back while I am asleep." Mentions of this kind of "external condition" are interpreted in the California studies as projections of neurotic anxiety into some irrelevant situations; this interpretation is much less applicable to the responses of the Russian group which, in terms of the California variables, seem rather to indicate concern with "threatening or non-supporting environment."

The next largest category for the Russians is that of "interpersonal situations" (24 Russians, 16 Americans*). It comprises two distinct subcategories which show a clear-cut differentiation between the two groups. The majority of the Russian responses refer to personal disharmony as an *interactive relationship* in which the subject feels himself participating, be it in the active or in the passive role (23 Russians, 6 Americans**). The responses of the Americans refer more often to moods and traits of *other people* (grouchy people, overbearing people, irresponsible drivers), with the subject himself as an outsider, suffering from these offensive traits (2 Russians, 11 Americans**).

In terms of California variables the Russian responses may be viewed as indicating "libidinized interpersonal relationships." This interpretation is also supported by the frequency with which the family is mentioned, whether in terms of longing, of concern for their well-being, or of being upset by discord (12 Russians, 1 American**). Furthermore the Russian responses also contain references to injustice and to the sufferings of other people, particularly of the Russian people under the Soviet regime, and to the lack of understanding of their situation by the West: "When one does not understand the Russians, when one confuses Russia with the Soviet Union—I will argue till I drop unconscious that the Russian people are not, as a whole, infected with Communism." These semi-ideological responses based on identification with a group have no counterpart in the American discussions (13 Russians**).

The category most frequently used by the Americans is that of "own moods." While the Russians use it with almost equal frequency (17 Russians, 21 Americans), the specific content of the responses of the two groups shows as much difference as in the case of "external conditions" and of "interpersonal situations." The Russians talk about longing, loneliness, fear, and moral sufferings; they never mention boredom, irritability, and restless tension which function prominently in the American responses. These objectless moods form a transition to the category of "bodily conditions" which is used by the Americans only and which contains complaints about poor health, sleeplessness, and—most frequently—fatigue (10 Americans**). In terms of California variables such references to bodily conditions are assumed to express "ego-alien trends," such as passivity, anxiety, and hostility. On the other hand, the Americans also exceed the Russians, though not significantly so, in the category "frustration of achievement" (10 Americans, 5 Russians), a type of response which is supposed to reflect conscious conflict and self-criticism.

Question 2. All of us have some desires which we try to suppress. Which desires and feelings do you find most difficult to control, most difficult to suppress?

The Russians made some references to specific conditions of their present and past, talking, e.g., about how difficult it was to suppress the wish to fight, to protest against the Soviet regime, or to banish painful thoughts about the relatives left behind. However, such references were much less frequent than in Question 1. Apart from this special category and from the Russians' greater difficulty in understanding the question, the distribution of the major content was fairly similar in the two groups.

In each group one third of the subjects made references to difficulties in the "control of primitive impulses." Within this category the Russians mentioned aggression and drinking with equal frequency, while for the Americans the concern with hostile impulses pre-

dominated over all others (7 Russians, 13 Americans). In terms of the California variables this might mean the prevalence of "nonfocal aggression" in the American responses, and a greater frequency of the "incidental pleasures" variable in those of the Russians. However, the difference between the two groups was significant only in the subcategory of drinking, which was used only by the Russians (7 Russians*).

The Americans used two categories that were absent in the Russian material. The first comprised the wishes "to quit," "to go off," "to do what I want to do and not what I must": to get away from the obligatory routine (7 Americans*). Such responses are interpreted in the California studies as indicative of "ego-alien passivity." The second category referred to the difficulty the subjects had in suppressing other disturbing feelings and moods (5 Americans*). The third minor category, which comprised "wishes for fame and material possessions," showed no differentiation between the two groups (7 Russians, 7 Americans).

Question 3. Whom do you consider to be really great people? What kind of people do you admire?

The bulk of the responses of both groups consisted of names of admired people, given with or without elaboration. They fall into two major categories: "artists and scientists" (named by 34 Russians and 28 Americans), and "social leaders" (23 Russians, 28 Americans). A smaller proportion of responses of both groups was given in terms of admired personal traits.

The subcategories of "social leaders" yielded no significant differences between the two groups, though the Americans tended to favor liberal leaders, such as Lincoln and F. D. Roosevelt (13 Russians, 20 Americans); the Russians, while also paying tribute to those "who worked and lived for the good of the people," gave no less attention to names that connote national power, including the military (15 Russians, 8 Americans). It should be noted, however, that in the latter subcategory the Russian list of choices was headed by Peter the Great, who represents progress and reform as much as national power; among the purely military leaders the Russians gave prominence to Generl Kutuzov who is credited with defeating Napoleon and who stands for victorious defense of the motherland rather than for conquest.

Pronounced differences between the two groups appear in the subcategories of "artists and scientists." Scientists are mentioned by them with about equal frequency (7 Russians, 10 Americans), but while the Americans give only a slightly greater prominence to artists of all descriptions, the Russians mention the latter with much greater frequency (33 Russians, 13 Americans**). When the artists are subdivided into writers (including poets), musicians, and painters, the Americans are shown to distribute their choices among these groups evenly, while the Russians favor the writers (31 Russians, 7 Americans**), naming musicians much less frequently and painters hardly at all. The difference between the two groups with regard to admiration for writers appears even more striking if we take into account that the Russians gave 86 single responses in this category and mentioned 34 different names, while the corresponding frequencies for the Americans are 10 and 8. Prominent among the writers and poets mentioned by the Russians were Pushkin, Tolstoi, Dostoevsky, Lermontov, Yessenin; of the foreign writers Shakespeare, Victor Hugo, Jack London, and Mark Twain were mentioned more than once.

The Russians' great admiration for writers is a continued cultural tradition; it may indicate the high value they place on expression and representation of human emotions, as compared with objective understanding and with practical action. The greater stress placed on these latter by Americans is reflected in the relatively greater weight they give to science and to social leadership, in their occasionally naming businessmen and sportsmen whom the Russians do not mention at all, and in the nature of the character qualities they consider admirable. Within this latter category they give primacy to traits related to practical achievement, such as strength, capability, success, while the Russians stress more inward and "moral" attributes, such as sincerity and kindness. However, the only difference within these minor categories that reaches statistical significance is the greater frequency with which

American subjects express admiration for their own relatives (2 Russians, 9 Americans*).

Question 4. Nearly every person says at times to himself: If this goes on I shall go out of my mind. What things can make people go out of their minds, lose their senses?

In this question the frequency rank order of the three major categories is the same for both groups, although the actual frequencies and the specific content of responses are different in some cases. For both groups the main cause of mental disturbance lies in internal psychological conditions, in "emotional and mental stress" (26 Russians, 30 Americans). In the category "external conditions" the Russians predominate over the Americans (20 Russians, 13 Americans), though not significantly so. This relationship is reversed in "disturbed personal relations" (4 Russians, 14 Americans*).

The differences between the responses of the two groups which appear in the subcategories are similar to those observed in Question 1. Of the Russian responses coded under "external conditions," three-fourths refer to conditions of extreme pressures and threats which are not mentioned by the Americans (14 Russians**); on the other hand, the Russians never mention minor environmental disturbances such as noises, disorder, traffic jams, referred to in more than half of the American responses (7 Americans**). The few Russian remarks coded under "disturbed personal relations" refer to interpersonal discord, while the Americans take a more passive attitude in their mentions of being persecuted, criticized, exposed to a nagging wife or to a drunken husband, or of being isolated from "normal relationships." It is interesting to note that while in Question 1 the Russians stressed the importance of personal relations for their happiness or unhappiness more than did the Americans, in thinking about the causes of a more catastrophic breakdown the latter show more concern than the former about the disastrous consequences of conflict and isolation.

In the leading category of "emotional and mental stress," the Russians talk mostly about stirring traumatic experiences which may overpower the person, such as terror, deep grief, or unhappy love (26 Russians, 11 Americans**); only the Russians occasionally blame insanity on lack of control of emotions or of impulses (5 Russians*). The Americans ascribe mental illness predominantly to conflicts and frustrations created by internal obstacles, by "blocks within onself," "no outlets for your emotions": a type of explanation never given by the Russians (20 Americans**). This difference in the concept of mental stress can be traced in part to linguistic factors and to relatively superficial cultural influences. Since psychoanalytically oriented theories have been condemned by the "party line," even the educated Russians have not been exposed to them as the American public has been. Terms like "repression" or "defenses" are absent from their vocabulary and from their thinking; the popular term "frustration" which was the one most frequently used by the Americans in the category of emotional stress has no exact counterpart in Russian. However, the difference in the conception of emotional causes of insanity might also reflect some real differences of psychological functioning. The apprehension of the Russians that they will be disorganized by overstrong emotions might be a consequence of their greater emotional abandon and impulsiveness; such experiences could make them stress the necessity of deliberate control, of holding impulses and emotions in check. In terms of California variables, most of their responses would fall under the category of "too much inner life." The Americans, on the other hand, sound as if they were greatly hemmed in by automatic defenses against strong emotional experiences to the extent that they possess an introspective orientation, and, in the vocabulary for describing these "inner psychological states" of conflict and tension, they express their apprehension of the consequences of this situation by ascribing insanity to frustration and repression.

Question 5. What do you think is the worst crime a person can commit?

The main content category which comprises one-half of the responses of each group is "murder." This category was subdivided, as in the California studies, into plain unqualified references to murder, and into responses that, by bringing in the motives for the act, or by some other qualifications and

elaborations, go beyond the reference to the external act as such. The first subcategory showed no difference between the two groups (20 Russians, 20 Americans); however, more Russians than Americans gave responses in the qualified subcategory, such as "murder committed for financial gain," or "murder by the government" (16 Russians, 5 Americans**).

References to stealing.and deceit were made by Russian subjects only (9 Russians**). References to "offenses against oneself and one's values" were made only by Americans (5 Americans*). Betrayal of friends or country, cruelty, and sexual offenses were mentioned by both groups with equally low frequency.

Question 6. All of us get sometimes into situations when we feel very much ashamed and embarrassed. What experiences make you feel like sinking through the floor?

The bulk (three-fourths) of each group's responses to this question fall into three categories: "violation of moral values," "inadequacy," and "violation of conventions." The first refers to failure to live up to one's obligations, promises, or to moral standards of behavior; the formulation often implies that the person was ashamed of what he had done, and not merely humiliated by exposure. The second category refers to experiences of personal inadequacy and inability, either in work or in interpersonal situations. Responses in the third category show a preoccupation with conventional rules of behavior and etiquette and a fear of appearing ridiculous: having one's clothes in disorder while in public, making an involuntary *faux pas* in using a foreign language. This category may be considered as the opposite of the first, because the emphasis here is shifted from essential values to superficies of behavior, and from one's own concern with one's action to the unfavorable reaction of others. The category of "inadequacy" occupies a middle positior between the two. Some responses in this category reflect the person's own concern with achievement and personal adequacy, but more often than in the case of "moral values" the cause of painful embarrassment lies in the exposure of one's inadequacies to others.

The distribution of responses over these categories is very different for the two groups. While the Americans use the three categories with approximately equal frequency, the Russians refer most frequently to moral values (30 Russians, 13 Americans**), and very rarely to conventions (6 Russians, 16 Americans*). This distribution suggests that the Russians are less afraid of exposure of their personal weaknesses to others, and experience less social anxiety than do the Americans. No difference was found between the two groups in the frequency of the use of the category "inadequacy" (13 Russians, 17 Americans), but its specific content tends to be different. The Americans talk almost exclusively about intellectual inferiority, mistakes, or failure to work, being proven wrong or criticized in front of others; the DPs' feelings of personal inadequacy are often aroused by having no work and no earnings, by being forced to ask for help. In "violation of moral values" both groups equally emphasize lying and dishonesty, with a minor focus on "hurting people" in other ways, but while the Americans express these concerns in a rather impersonal or self-contained way ("failure to live up to promises," "to discover I was offensive"), the Russians often make the situation concrete by referring to specific acts and to particular people involved; they feel regret, e.g., if they had let down a friend, or had been rude to an old person. As in Question 5, occasional references to stealing are made only by the Russian subjects.

Question 7. Suppose you knew you had only six months to live, but during that time you could do whatever you liked—how would you spend your time?

This question invites the subjects to explore in what direction his wishes would take him if all obstacles to his actions, including their long-term consequences, were removed. The question also evokes reactions to imminent death. In both respects there are pronounced differences between the two groups.

References to "enjoyment, pleasure" form the main category for both groups (28 Russians, 29 Americans); however, the distribution of responses within this category is quite different. In the Russian group references to frankly *sensual pleasures*, such as eating, drinking, and sex, outweigh almost two to one the mentions of such pleasures as travel, enjoyment of arts and sports (sensual pleasures: 22 Russians, 8 Americans**). In terms of

California categories the predominant Russian responses belong under "open sensuality and active pleasure." Conversely among the Americans, references to what one subject termed *"the finer things of life"* predominate over the cruder pleasures in the relationship of three to one, "travel" being their most frequent single response (11 Russians, 25 Americans**). These responses have a marked affinity to the "incidental, dilute pleasures" of the California classification.

Other differences are less pronounced. The Russians give more responses in the category of "religious and moral preoccupations," expressing these concerns either in pure form or as an integral part of some conflict about pleasure seeking or impulse gratification (16 Russians, 6 Americans*). They also tend to dwell more on the "affiliative ties," if this category is made to include not only the wish to spend time with family and friends, but also the wish to see Russia once more (14 Russians, 9 Americans). Neither group expresses any desire to utilize the time left for contributing to social welfare, to mankind in general; however, some of the Russians wish to relieve the suffering of the Russian people, e.g., by committing a decisive terroristic act, such as killing Stalin. The Americans emphasize personal achievement slightly more than do the Russians, particularly in the form of a wish to continue and complete some piece of work.

In expressing their reaction to imminent death, subjects in both groups comment that this knowledge might deprive them of any enjoyment and enterprise. However, while the Russians occasionally talk about "going crazy" or "sitting and crying," of committing suicide, the Americans often deny the emotional upset explicitly or by implication, maintaining that they would carry on as usual; some of them express the wish to keep their fate secret from their families and friends.

Question 8. We have the feeling of awe when something seems to us remarkable, beautiful, or important. What things or events inspire such feelings in you?

Even though an elucidation of the word "awe" was included in the question, this term was frequently not understood, or misunderstood, particularly by the Russians. They often gave it the meaning of "admiration" and occasionally the meaning of "pleasure" or "delight." Responses referring to such events as being given a chance to emigrate are caused by this misunderstanding which possibly also accounts for some of the differences in the responses of the two groups.

In the American group the leading category is that of "individual achievement" (10 Russians, 28 Americans**). In more than half of the instances the Americans specify this achievement as scientific or artistic, but—unlike the Russians—they also make references to outstanding performance in technology and in sports (1 Russian, 10 Americans*), while the Russians refer to arts and sciences almost exclusively. The second place for the Americans is occupied by "nature," which includes also references to life processes (6 Russians, 16 Americans*). In the Russian group the leading category is the "social-interpersonal": the subjects feel moved by warm personal relations and altruistic acts, as well as by events of wider social significance such as the triumph of freedom (15 Russians, 7 Americans*). Responses expressing "national pride" were equally infrequent in both groups (5 Russians, 7 Americans).

Within the categories used by both groups, the formulations of the Americans tend to be more schematic and conventional than those of the Russians, indicating a less intense and personal mode of experience. Thus in the category of "nature and life" most of the American responses are so brief and stereotyped that they must be described as "dilute experiences" in terms of the California categories. The Russian responses are much more concrete and detailed, more suffused with personal meaning and emotion, and consequently may be considered as representing "intense nature experiences."

Question 9. Suppose you had a child, and you knew that you might wish three things for him and that your wishes would be fulfilled—what would you wish for?[6]

Apart from the fact that the Americans gave many more general unspecified responses, such as "happiness," (5 Russians, 22 Americans**), the wishes for the child were fairly similar in the two groups. The largest category

[6] This question has been used in T. Dembo's study, "Investigation of Concrete Value Systems," U. S. Public Health Service, Institute of Mental Health. Final Report (Unpublished).

contained wishes for a "successful life" or for a life lived under favorable conditions (30 Russians, 34 Americans); in this category the Russians stressed particularly the importance of education and of specialized training as prerequisites for the "good life" (17 Russians, 9 Americans). The Americans, on the other hand, placed a greater emphasis on wealth, on financial status (8 Russians, 16 Americans*).

While both groups wished for good health for the child (20 Russians, 27 Americans), the Russians emphasized no less strongly mental qualities (18 Russians, 12 Americans) and moral or character traits (23 Russians, 11 Americans**) to which the Americans paid less attention. In speaking of mental qualities, the latter mentioned intelligence almost exclusively; the Russians divided their wishes between intelligence and special talents, particularly the artistic ones (9 Russians, 2 Americans*). The Americans showed slightly more concern than the Russians with the child's social relations and also made occasional references to "full life," to "secure and harmonious development," which had no counterpart in the Russian responses. The Russians occasionally expressed the wish that the child should grow up to love Russia, and be able to apply his talents in Russia.

DISCUSSION

In reviewing the trends revealed in the responses to the nine questions, we shall first organize the discussion in terms of the positive and negative values held by the two groups, and then consider the implications of our findings for some aspects of the theory of the "authoritarian personality." We shall use the term "value" in a wide sense, including not only explicit evaluations, but also consistently positive or negative emotional reactions of a more immediate kind, as they are expressed both in the content and in the formulation of responses. It is well to remind the reader at this point that the generalizations that follow are based merely on the differential frequencies of responses in some of the content categories and that many of these differences are small, or occur within minor, infrequently used categories. In several questions there is a considerable similarity between the two groups, at least in the distribution of the major content categories. Furthermore, some of the differences that do appear are obviously created or enhanced by the drastic differences in their life situations both in the present and in the recent past, and cannot be ascribed exclusively to differences in permanently held values. On the other side of the ledger, it is to be noted that many of the differences observed are not explicable in situational terms alone, and that several of the intergroup differences appear quite consistently from question to question, even though they may reach the level of statistical significance only in some of them.

In reviewing those questions that elicit expressions of *positive values* (Questions 3, 7, 8, 9), we find that two major values are more prominent in the Russian records than in the American ones. The first one is the value of emotional experiences and of frank sensual pleasures. The Russians' positive acceptance of these experiences is expressed both in the content of their responses and in the language they use, which is more emotionally expressive and more concretely descriptive of sensual impressions than that used by the American subjects. Even more pronounced is the Russians' high valuation of interpersonal relationships, their positive and active acceptance of interaction and of belongingness with others. This attitude is expressed in most of the questions, both in an immediate emotional fashion and on an ideological level. The other two features more prominent in the Russian than in the American responses—moral values and patriotic feelings—also have strong interpersonal implications for our subjects. Patriotism is often expressed in the context of wishing to help the Russian people; moral values, such as sincerity and honesty, are seen as prerequisites of friendship.

The value which the Russians emphasize less strongly than do the Americans is first and foremost that of individual achievement; the latter mention it more frequently than do the former in practically all contexts that evoke discussions of this topic. Within this category what seems distinctively American is the high valuation of rational knowledge and of organizational and physical achievement. A second positive category in which the Americans have the advantage is an infrequent one but it has practically no Russian representation. This category includes responses emphasizing independence and self-expression of the

individual and a protest against their violation; the concept of "crime against oneself" (e.g., "deceiving oneself") also belongs to this category, which could be termed "integrity of the individual."

In response to inquiries about *negative values* and sources of negative experiences (Questions 1, 2, 4, 5, 6), the Russians much more often than the Americans see causes of unhappiness in severely depriving conditions of life, conditions that obstruct satisfaction of vital needs or threaten one's life or safety. The other two categories in which their responses by and large exceed those of the Americans refer to disturbance of interpersonal relationships and to violation of moral values. These three categories, which can be viewed as the obverse of the Russian positive values, belong to three different spheres: external, interpersonal, and ethical. Yet all of these categories refer to vital conditions of human existence and imply the person's essential interrelatedness with his environment. Unhappiness is seen by the Russians as resulting from the disruption of this relationship either by the "environment" (depriving conditions) or by the person (moral violation), or by both (interpersonal discord).

In contrast to this relative homogeneity, the negative conditions that are mentioned by the comparison group more often than by the Russians fall into two major categories which seem to be quite different in their psychological meaning. On the one hand, consistent with their positive values of individual achievement and individual integrity, the Americans locate causes of unhappiness within the person himself: in his conflicts, inhibitions, in his feelings of inadequacy and failure. These responses presuppose a self-reflective, self-critical attitude which is less prominent among the Russians. On the other hand, we find numerous references to inessential, often trivial conditions which have little relation to basic human needs, and which are located "outside": in other people, in things, in the body. Annoyances produced by noises, bossy people, insomnia, breaches of etiquette, or impulsive wishes to "go off" are likely to be merely symptoms of disturbances caused by unconscious conflicts. Such responses indicate a defensive alienation from one's feelings which is the opposite of insight. Yet both categories have the common feature of implying the person's separateness rather than close relatedness with his environment.

Thus a review of the distinctive values of the two groups suggests that the Russians are more firmly and securely integrated with their environment than are the Americans, who are more keenly aware of the individual's separateness and isolation. Because the most significant part of our environment is people, it is understandable that the most pronounced differences between the two groups appear in the interpersonal area.

Considering our findings in terms of the subvariables of the equalitarian and authoritarian personality patterns, we arrive at the following formulations. The equalitarian patterns of the Russians are based to a greater extent than those of the Americans on the area of *interpersonal relationships*, while those of the Americans stem from their concern with the *rights and achievements of the individual*. Within the area of inner experience, the Russians' equalitarian personality structure is attested to by their *acceptance of impulses and emotions*, and by moral self-reproach, while the corresponding American pattern finds expression in a more rational and self-critical *introspective attitude* and in feelings of inadequacy and rejection. The authoritarian patterns of the two groups also show differences but along different lines. If one were to score all Russian responses according to the specific rules of the California study which have been worked out for the American groups, a greater proportion of the Russian than of the American "high" scores would be earned through "moralizing" discussions and through expressions of patriotic sentiments—i.e., through their *ideological attitudes*. The high scores of the Americans are more often founded on *defensive personality aspects*, such as alienation from oneself and from others and displacement of emotions into irrelevant situations.

These generalizations must be qualified by some further observations which are pertinent to the status of ideological attitudes as indicators of personality patterns. The wording and the context of our subjects' responses suggest that certain explicit evaluations do not have the same significance for the two national groups, and this impression is confirmed by their differential correlations with the over-all

scores of authoritarianism. Thus the Russians' traditional admiration for great writers does not seem to be diagnostic of personal equalitarianism to the same degree as it is for the Americans, nor do their feelings for their country appear to indicate ethnocentric authoritarian patriotism. Insofar as the choice of particular values and opinions, even when guided by personal motives, depends also on their prevalence and meaning in the cultural environment, such differences are to be expected. The more direct dependence of ideological attitudes on the environment makes them less valid as indicators of either "high" or "low" patterns than are the more deeply ingrained personal traits, particularly when subjects of different cultural backgrounds are being compared. Since our delineation of the difference in the authoritarian patterns of the two groups is based in part on the Russians' ideological pronouncements, it appears less valid than the generalizations that concern the intergroup differences in equalitarianism.[7]

By comparison with the attitudes of Americans, the equalitarian traits of the Russians appear to be more pronounced in the interpersonal than in the "individual" area, and in the emotional rather than in the intellectual sphere. What implications do these findings have for the general theory of authoritarianism? Primarily they serve to confirm the existence of subvariants, at least within the equalitarian pattern, and to identify the two variants that are distinctive of the two national groups: one centered on interpersonal relatedness and immediate emotional awareness, the other on the integrity of the individual and on rational self-reflection. Though these two patterns reflect different value-orientations, and possibly also different levels of sophistication, within the theoretical framework of the California studies they may be considered as equally expressive of genuine equalitarianism. It is a task for future studies to verify or disprove the existence of these and of other possible patterns by a systematic investigation of the correlations between the many specific traits that are considered as diagnostic of equalitarianism and of authoritarianism.

SUMMARY

The projective questions used in the California studies of authoritarianism were administered to a group of former Soviet citizens and to a comparison group of Americans. The responses were coded according to their content. The findings are discussed in terms of the emotional and evaluative attitudes prevalent in each group, and their implications

[7] The formulations concerning the equalitarian patterns are also borne out by the evidence of some other tests we used in the study. In comparing the subjects' scores on different tests J. Orton found this evidence to be particularly clear-cut on the point of "interpersonal" vs. individual approach. One of the items in the Episodes Test (2) depicted a conflict between the individual and the group of which he is a member. In discussing this situation, the Russians in general identified with the group more strongly than with the individual, but this tendency was even stronger among those whose scores of authoritarianism on the Projective Questions were low than among the "highs." The Americans as a group displayed a stronger identification with the individual, and here the "lows" did so to a greater extent than the "highs." To understand this paradox, one must take into account that the Russians identified with the group in a very positive, participating way, while the majority of the Americans described the group as a coercive agent to be either fought with or yielded to, i.e., in terms of an irrational authority.

Some evidence on the "emotional" aspect of the "low" Russian pattern was obtained from the technique of "short answer items," used with the Russian group only. The responses of the Russians to these questions, most of which pertained to various areas of Soviet life, were scored in terms of the four scales of Flattery, Distortion, Rejection of the basic Soviet institutions, and general Anti-Soviet sentiment. Flattery (i.e., opinions favoring the Americans as, for example, over the British) was higher among the "highs" than among the "lows": the former are more eager to please the momentary authority. For the rest of the variables the relationship was reversed. The "lows" not only expressed a much stronger anti-Soviet sentiment than the "highs," but also went much further in rejecting such widely accepted institutions as the state's ownership of industry, and even in denying such recognized achievements of the regime as increase in literacy and in production of farm equipment (Distortion). It would seem that respect for either factual objectivity or for achievement as such is not a component of the Russian "low" pattern. Although these subjects would not falsify their evaluation of the American merits and failings while talking to us, for them to pass an objective judgment on the instrumental achievements of the Soviets would violate the emotional truth in a more important way than would detached objectivity. This interpretation is borne out by their frequent alternative reaction to inquiries about the Soviet achievements: "Yes, but at what price—and to what end!"

for the theory of authoritarian personality are pointed out.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. HANFMANN, EUGENIA, & GETZELS, J. W. Interpersonal attitudes of former Soviet citizens, as studied by a semi-projective method. *Psychol. Monogr.*, 1955, **69**, No. 4 (Whole No. 389).
3. ZUBIN, J. Nomographs for determining the significance of the differences between the frequencies of events in two contrasted series or groups. *J. Amer. statist. Ass.*, 1939, **34**, 539–544.

Received July 25, 1955.

VOLUNTEERING AS AN AVOIDANCE ACT

ROBERT R. BLAKE, HOWARD BERKOWITZ, ROY Q. BELLAMY
AND JANE SRYGLEY MOUTON

The University of Texas

VOLUNTEERING is a dynamic social process. The amount of volunteering can be increased or decreased by changing the structure of the force field acting on the person in the specific situation. Some of the factors determining the resultant force toward or away from accepting a volunteering request already are known. For example, volunteering can be increased by others accepting (8, 9) and decreased by others refusing the request (8). The strength of the request also influences acceptance since frequency varies directly with the compellingness of the induction (7). Still another factor is whether or not acceptance is in public or private, with more volunteering under private conditions of induction (9). Although clear relationships between personality and volunteering have not been demonstrated (6), it can be presumed that personal variance is a factor in the resultant force toward or away from volunteering.

The hypothesis tested in this study is that changes in the frequency of volunteering in response to a standard request are associated with variations in the attractiveness of possible alternative actions. The prediction is that the more attractive the alternative the lower the frequency, and, conversely, the lesser its attractiveness the higher the volunteering rate. The prediction is the same whether made on the basis of a Lewinian analysis (3), adaptation-level theory (1, 2), or from approach-avoidance concepts (4). Since a previous experiment has demonstrated differential acceptance rates for requests under public and private conditions (9), the hypothesis tested here is evaluated for both conditions of responding.

METHOD

Subjects and Arrangements

Subjects included 362 students in undergraduate classes at The University of Texas. Six classes served under the private condition. Slips were distributed which students were requested to sign; they were asked to check either: "I wish to participate in the experiment," or "I do not wish to participate in the experiment." Three additional classes served under the public condition where students indicated acceptance by raising their hands.

Variations in the Attractiveness to the Alternative to Volunteering Condition 1 (Attractiveness associated with not volunteering).

A positive expectancy for the alternative to volunteering was created by the instructor's informing the class he would be called out during the hour. He indicated that the class would be dismissed for the remainder of the period. The experimenter then entered to solicit volunteers. After a few seconds' hesitation, the instructor agreed to excuse from class those who volunteered.

Condition 2 (Control condition). Shortly after the instructor had begun to lecture, the experimenter entered and solicited volunteers as described above. The instructor agreed to excuse from the remainder of the lecture those who volunteered.

Condition 3 (Volunteering as an avoidance act). A negative expectancy was created by the instructor's announcing a "pop quiz." While the questions were being written on the blackboard, the requester entered to solicit volunteers. The instructor explained that a test was being given. After a few seconds' hesitation, he agreed to excuse from taking the test those who volunteered.

Request for Volunteers

After receiving permission to solicit volunteers, the requester made the following appeal. "We are conducting a psychological experiment and need some volunteers to participate this hour. It will take no longer than this class period." For the private condition he said, "Will each of you sign your name to the slips that are being pased around, indicating whether or not you will participate? Check one of the two statements on the slip." For the public condition the following statement was substituted for the last sentence, "Will those of you willing to participate raise your hands?" A full explanation was provided after completion of the study, since subjects were unaware that the arrangements were part of an experiment.

RESULTS

The data for the three conditions for both "private" and "public" volunteering are presented in Table 1. Consistent with prediction, the lowest rate was obtained when the alternative was attractive (Condition 1). Being dismissed produced significantly less volunteering than the opposite condition involving the unattractive alternative of a "pop quiz" and less, but not significantly so, than the control condition. The prediction of a higher rate for Condition 3 also is confirmed since the unattractive choice of taking a "pop quiz" is associated

TABLE 1

FREQUENCY OF VOLUNTEERING AS A FUNCTION OF BACKGROUND CONDITIONS

Condition*	*N*	Number of Volunteers	Percentage by Condition
Private			
1	67	18	26.8
2	83	38	45.7
3	85	84	98.8
Public			
1	34	4	11.0
2	39	12	30.7
3	54	54	100.0

* In Condition 1, attractiveness is associated with not volunteering; in Condition 3, volunteering avoids an unpleasant situation; Condition 2 is neutral in these respects.

TABLE 2

CHI-SQUARE VALUES FOR VOLUNTEERING UNDER DIFFERENT BACKGROUND CONDITIONS

Condition*		χ^2	*df*	*p*
Private	1 × 2 × 3	22.83	2	.01
	1 × 2	2.53	1	—
	2 × 3	9.82	1	.01
	1 × 3	18.83	1	.01
Public	1 × 2 × 3	24.80	2	.01
	1 × 2	2.79	1	—
	2 × 3	9.65	1	.01
	1 × 3	16.36	2	.01
Private × Public		.54	1	—

* In Condition 1, attractiveness is associated with not volunteering; in Condition 3, volunteering avoids an unpleasant situation; Condition 2 is neutral in these rsepects.

with significantly more volunteering than is found for the control condition or for Condition 1. Here volunteering is an avoidance act. The invitation is accepted in order to leave an undesirable situation. Frequency for the control condition gives a rate intermediate between the extremes. The significance of the differences between conditions are shown in the χ^2s in Table 2.

The same relationships were obtained for both public and private conditions, though volunteering in private for Conditions 1 and 2, is greater than volunteering in public. The latter trend though not significant is similar to the one reported by Schacter and Hall (9). The findings gain further support from a study where personal independence in the Asch-type situation was maintained more frequently when the anonymity of the respondent was insured (5).

DISCUSSION

The rate of volunteering in any concrete situation, then, is a function of the social field structure. The strength and direction of the resultant force determines the frequency with which acceptance occurs. The attractiveness of the requested action relative to that of available alternatives is a critical determinant of the resultant force. The general principle is that the rate can be changed by increasing or decreasing the desirability of the task, by increasing or decreasing either the attractiveness of alternative actions, or by both in varying proportions.

The findings can be evaluated from the standpoint that the subjects reported experiencing the decision to volunteer as a matter of personal preference. Participants saw little or no connection between the experimentally created expectancies for the class hour and the request for volunteers. They felt they were acting on personal preference, yet from the standpoint of the experimental design, volunteering was regulated by the relative attractiveness of the experimentally created alternatives. The difference between the situation experienced by the subject and that which occurred provides a basis for assigning meaning to the term "volunteering." The subject feeling that action is volunteered, or governed solely by personal preference, occurs whenever the respondent sees no causal relationship between the various factors that are involved in his decision. He feels he has made a personal decision concerning participation, yet from a systematic point the resultant force that determines his decision is based on a predictable weighting of all relevant factors operating in the situation. Thus, decisions that in personal experience seem to be free, independent, and personal are, from another standpoint, determined by the psychological properties of the force field within which the action occurs.

Examples may be given applying the findings reported here. Why do people volunteer for work such as in the American Red Cross, the American Cancer Society or the Girl Scouts of America? The decision is determined by the attractiveness of the work, relative to possible alternative activities. Military recruitment through voluntary enlistment also can be viewed in the same manner. The rate of enlistment is high when civilian employment is diffi-

cult to procure, as in times of depression, and low during prosperous times, when alternative civilian employment is more desirable.

SUMMARY

Volunteering as a function of the relation between the attractiveness of the requested action and that of the alternative to volunteering was studied in the present experiment. Three hundred sixty-two subjects reacted under public or private conditions to a standard volunteering request. Attractiveness was varied so that under one set of conditions volunteering would be more desirable than the alternative, while under another it would be relatively more undesirable. Under the control condition attractiveness of the alternative was not altered. Results support the predictions. Volunteering was significantly more frequent when the alternative was less attractive than the act and less frequent when the alternative was more attractive than the requested action. Frequency under the control condition was intermediate between the other two. The conception of volunteering as a dynamic social process resulting from the structure of the force field was discussed.

REFERENCES

1. BLAKE, R. R., & MOUTON, JANE S. The study of social conduct within the framework of adaptation-level theory. In M. Sherif & M. O. Wilson (Ed.) *Emerging problems in social psychology.* Third conf. in Social Psychol. Norman: Univer. of Oklahoma, in press.
2. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
3. LEWIN, K. *Field theory in social science.* New York: Harper, 1951.
4. MILLER, N. E. Comments on theoretical models illustrated by the development of a theory of conflict. *J. Pers.*, 1951, **20**, 82–100.
5. MOUTON, JANE S., BLAKE, R. R., & OLMSTEAD, J. A. The relationship between yielding and disclosing personality identity. *J. Pers.*, 1956, **24**, 339–347.
6. ROSEN, E. Differences between volunteers and non-volunteers for psychological studies. *J. appl. Psychol.*, 1951, **35**, 185–193.
7. ROSENBAUM, M. The effect of stimulus and background factors on the volunteering response. Unpublished doctor's dissertation, Univer. of Texas, 1954.
8. ROSENBAUM, M., & BLAKE, R. R. Volunteering as a function of field structure. *J. abnorm. soc. Psychol.*, 1955, **50**, 193–196.
9. SCHACHTER, S., & HALL, R. Group-derived restraints and audience persuasion. *Hum. Relat.*, 1952, **5**, 397–406.

Received September 22, 1955.

THE STRUCTURE OF SOCIAL ATTITUDES IN LEBANON

E. TERRY PROTHRO AND JACK D. KEEHN
American University of Beirut

ONE of the authors of *The Authoritarian Personality* has observed (12, p. 73) that in spite of the enormous number of studies inspired by publication of that volume, the identification of the authoritarian syndrome is still far from established. In examining data on this subject, Christie (3, p. 140) was reminded of Augustine's comment about time: "For so it is, oh Lord my God, I measure it, but what it is that I measure I do not know."

In contrast to the uncertainty of these writers is the opinion of Eysenck (5, p. 152), who states quite simply that "the F-scale is essentially a measure of tough-mindedness." By "tough-mindedness" he means performance on a scale developed by himself (4) and co-workers (5, p. 276). The T scale is unifactorial and some idea of its importance may be gained from the assertion that it, together with the unifactorial R scale designed to measure radicalism-conservatism, determine almost all correlations between all attitudes in England, the United States, Sweden, Germany, and other countries having similar forms of social organization (5, pp. 265–266). If Eysenck's view can be accepted, then it becomes clear that social psychologists have been neglecting the key to many of their most difficult problems—including that of the authoritarian syndrome.

The evidence cited for the proposition that "the F-scale is essentially a measure of tough-mindedness" is far from conclusive. Coulter, in an unpublished doctoral dissertation (5, pp. 152–153) found a correlation of .43 between the T scale and the F scale for a group of 83 soldiers, a correlation of .62 for a group of "male Fascists," and a correlation of .63 for a group of "male Communists." The probitive value of this evidence should be clear to anyone familiar with the vast number of scales and indices which have moderate positive correlations with the F scale (10, p. 23).

Coulter also found that her Communists scored higher on the F scale than did her unselected soldiers, and Eysenck seems to feel that this reveals new information about the F scale. This datum is sufficiently surprising to call for closer examination, for other evidence (11, pp. 24 ff.; 3, p. 131 ff.) points in the opposite direction, but in any event it bears only indirectly on the identity of F and T. Consequently it appears that the suggestion that the F scale and the T scale are measuring essentially the same thing must be viewed as a bold but unproved assertion.

Keehn (7) administered Eysenck's public-opinion inventory (4) in Lebanon with a view to determining whether the R and T dimensions would emerge. He did find two orthogonal factors which he labeled "Arab nationalism" and "toughmindedness-tendermindedness." He concluded that his results were compatible with those of Eysenck, but that Eysenck's R factor was a special case of the general dimension which might be called "the major political issue of the culture under investigation." It should be noted, however, that for most items Keehn's two factors account for less than 20 per cent of the total item variance.

It has been shown that the F scale is a useful and meaningful instrument for measuring authoritarianism in Lebanon (9) in spite of cultural differences between California and Lebanon. Some question was raised as to the meaning of authoritarianism, however, when it was found that there was no correlation in Lebanon between the F scale and the California scale of politico-economic conservatism. Moreover, a follow-up study (8) revealed that there was no correlation between authoritarianism as shown by scores on the F scale and ethnocentrism as shown by performance on scales of social distance. In interpreting this lack of relationship it must be noted that there is no evidence that either the PEC scale or the social-distance scale taps a meaningful dimension of social attitudes in the Near East.

To explore the meaning of authoritarianism in such a setting it would seem appropriate to employ items which have emerged from factorial studies performed in the culture. Performance of Lebanese subjects on the F scale was therefore compared with their response to items representing each of the two factors found by Keehn, toughmindedness-tendermindedness and "Arab nationalism."

TABLE 1

FACTOR SATURATIONS OF ITEMS SUBJECTED TO ANALYSIS

Item*	Unrotated			Rotated		
	F_I	F_{II}	F_{III}	F_I''	F_{II}''	F_{III}''
F6	51	42	−35	73	06	−10
F8	−05	26	35	00	−07	42
F9	37	−32	12	−03	49	−05
F12	07	11	12	08	03	15
F13	78	−10	−15	46	61	−17
F16	45	18	−05	42	23	05
F18	59	37	05	58	27	23
F19	70	−28	12	19	74	−03
F21	35	36	−36	60	−03	−13
F25	40	16	12	30	26	18
F26	41	−05	−32	34	24	−30
F29	23	17	−30	38	−01	−17
F31	20	34	−15	41	−07	03
F33	06	−23	−57	11	00	−60
F34	54	−29	−13	18	53	−25
F35	25	37	13	36	04	29
F38	44	27	40	30	32	47
F39	47	−23	17	06	54	03
F42	34	50	26	45	08	46
F44	68	−23	42	10	78	25
T9	29	−13	14	−06	22	24
T19	22	26	−15	−36	02	00
T20	36	22	34	24	26	40
T36	−17	13	16	−08	−17	20
N8	−21	45	−26	27	−49	00
N13	33	−25	07	00	41	−07
N18	−54	54	−30	16	−80	00
N39	40	−26	15	01	48	00

* Both here and in the text the item numbers are those of the sources (1, pp. 255–257; 4, p. 566) rather than of our questionnaire. The letters F, T, and N refer to fascism, toughmindedness-tendermindedness, and nationalism respectively. See footnote 2 re item N13.

METHOD

Subjects were 100 students in General Education and Educational Psychology courses at the American University of Beirut. Of these, 72 were male and 28 female. With respect to nationality, 54 were Lebanese and 46 from nearby Arab states. During regular class periods they were given Eysenck's public-opinion inventory as adapted by Keehn (7) for use in the Near East together with an F scale of 27 items.[1] The format and instructions for the two scales followed the standard form of the original scales. Subjects worked anonymously. They had no difficulty in completing the two inventories in a single class session.

A centroid analysis was then conducted on items from the F scale and the Eysenck scale. The item intercorrelations for the analysis were tetrachoric correlations computed by use of the diagrams of Chesire *et al.* (2). Responses were dichotomized into "Agree" or "Disagree," with "Neutral" responses grouped with the smaller category in each instance. Because of the unreliability of such correlations when the dichotomy is near the tail of the distribution, items were omitted from the analysis when fewer than 20 per cent or more than 80 per cent of the subjects agreed with the item. Seven F-scale items were omitted for this reason. The factor analysis was performed on 28 items, of which 20 were from the F scale, 4 were from Keehn's toughminded-tenderminded scale, and 4 from the "nationalism" scale. The toughminded-tenderminded and "nationalism" items were selected on a basis of their high loadings in Keehn's previous study (7).

[1] All items of Form 45 were used except for items number 22, 30, and 43 (1, page 260). Two of these items referred to events at the close of World War II and one was a terse folk saying.

RESULTS

Factor saturations for the 28 items are shown in Table 1. Orthogonal rotations toward a simple structure were carried out graphically. There appear to be two significant factors running through the items analyzed. Analysis was terminated after extraction of three factors because the third factor was not statistically significant according to the methods of Tucker and Humphrey as described by Fruchter (6, pp. 77 ff.), and because only one of the second factor residual correlations reached the .01 level of significance whereas about four would have been expected by chance.

All but two of the F-scale items have positive loadings on Factor I″. Thus the belief that there is a general factor in those items is supported by our data. At the same time, however, it should be noted that some of the items have small loadings on this factor and large loadings on a second factor. The items which seem to be the best measures of Factor I″, in order of size of loadings are:

F6 human nature being what it is, there will always be war and conflict

F21 young people sometimes get rebellious ideas, but . . .

F18 . . . people move around and mix together so much, a person has to protect himself against catching a disease or infection from them

F13 what youth needs most is strict discipline . . . and the will to work and fight . . .

F42 no . . . decent person could ever think of hurting a . . . relative

F16 some people are born with an urge to jump from high places

F31 people are prying into matters that should remain personal . . .

F29 astrology can explain a lot of things

F35 sex life of the old Greeks and Romans was tame compared to . . . this country . . .

T19 major questions of national policy should be decided by reference to majority opinion (negative)

These items do not fit readily into any single cluster which has yet been found from analyses of the F scale. It seems to the writers that all of the items reflect what has been called a "low opinion of human nature" and of human effort. Item F42 might be interpreted as the traditional Near Eastern reliance on one's own in-group as the principal defense in a world where human nature is untrustworthy. We have therefore named this factor "cynicism," although admittedly other names and other interpretations are possible.

Factor II″ seems to be the factor which Keehn called "Arab nationalism." The four items which he used to define this dimension have significant loading on Factor II″. The items which our analysis shows to have loadings of .35 or more on this factor, in order of magnitude of loading, are:

N18 Jews are as valuable, honest and public-spirited citizens as any other group (negative)

F44 nobody ever learned anything . . . except through suffering

F19 an insult to our honor should always be punished

F13 what youth needs most is strict discipline . . . and the will to work and fight . . .

F39 homosexuals . . . ought to be severely punished

F34 most of our social problems would be solved if we could somehow get rid of the immoral, crooked and feebleminded people

F9 when a person has a problem . . . it is best for him not to think . . . but to keep busy . . .

N8 in the interests of peace, we must give up part of our national sovereignty (negative)

N39 the Japanese are by nature a cruel people

N13 refusal to fight for ones' country on religious grounds should be a punishable offense[2]

Some of these items might be thought of as characteristic of the views held by Arab nationalists, but items F39 and N39 would be hard to fit into such a concept. On the other hand, items with heavy loadings, such as F44, F19, F13, and F39 clearly reflect authoritarian aggression. It would be easier, then, to think of Factor II″ as one of authoritarian aggression or punitive outlook. Item N18 might appear on first reading to bespeak nothing but ethnocentrism. However, when the Arab-Israeli conflict and the problem of the Jewish minority in Lebanon are considered in connection with the item, it seems possible that punitiveness is involved in the response.

[2] This item was edited by Keehn (7) and the wording therefore differs from that which appeared in Eysenck's inventory (4, p. 566).

It is interesting to note that the authors of the F scale listed (1, p. 255) eight items which they had selected, on a priori grounds, as typifying authoritarian aggression. Six of these items were entered into our analysis, and four of them (F19, F13, F39, F34) had substantial loadings on Factor II″. A fifth item (F12) had virtually no loading, but it referred to "bad manners" and "breeding" and the item may simply be inappropriate in this culture. The sixth item (F25) had a loading of .26. The item referred to ". . . rape and attacks on children . . . such criminals ought to be publicly whipped, or worse." Such an item seems to reflect punitiveness in a high degree, and the low loading might raise some question about the identification of Factor II″. From discussion with local students and social scientists, the authors have discovered that public punishment is associated in the minds of many Arabs with the Ottoman occupation and with foreign mandatory powers. That is, such punishment is not thought of as an act on the part of the ingroup. Moreover, it has been pointed out to us that public punishment for sex crimes would make a public disgrace of an issue which might otherwise be settled by punitive action on the part of the family. It is possible, then, that these connotations of the phraseology of the item explain why it did not have a higher loading on the punitive factor.

Item F13 is listed as having a moderate loading on both cynicism and punitiveness. It is our interpretation that this item reflects a low opinion of the value of creative effort on the part of Arab youth and at the same time taps the punitive factor by its reference to fighting for family and country.

If we accept "punitiveness" rather than "nationalism" as the label for the second dimension, this implies that it is a dimension of personality and not of the political sphere. This point of view leaves open the possibility that some nationalists might hope to achieve national unity and independence without reliance on authoritarian aggression. It further implies that, contrary to Keehn's earlier suggestion, this factor cannot be linked with Eysenck's radical-conservative factor under the general heading of "major political issue of the culture under investigation."

The items which Keehn found to be measures of toughmindedness-tendermindedness in

Lebanon do not seem to be effective indicators of either Factor I″ or II″. The identification of authoritarianism with toughmindedness does not hold with our sample. Indeed, there is for this sample a correlation of only 0.19 between an F scale composed of the entire 27 items and a T scale composed of the 10 items which Keehn had found to be the best measures of T in Lebanon. For the F scale, the mean was 4.7 and the standard deviation .71. For the T scale, the mean was 1.86 and the standard deviation 5.11, where the highest possible score was 20 and the lowest possible score was −20.

SUMMARY

Previous studies with the F scale in the Near East had raised some question about the factors being measured by the scale. In order to gain more information on authoritarianism in this culture, an analysis was made of items from the F scale together with items from a scale of Eysenck's which seemed to measure toughmindedness and a political dimension. It was found that there was a general factor running through almost all of the F scale items, and this factor was tentatively labeled "cynicism." The "toughminded-tenderminded' items were poor measures of this factor. A second factor was extracted, which seemed to be the same as that previously identified as a political dimension. Our data suggested that this factor might better be interpreted on a personality dimension, which we tentatively labeled as "punitiveness" or "authoritarian aggression."

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. CHESHIRE, L., SAFFIR, M., & THURSTONE, L. L. *Computing diagrams for the tetrachoric correlation coefficient*. Chicago: Univer. of Chicago Bookstore, 1933.
3. CHRISTIE, R. Authoritarianism re-examined. In R. Christie & Marie Jahoda (Eds.). *Studies in the scope and method of "The Authoritarian Personality."* Glencoe, Ill.: Free Press, 1954. Pp. 123–196.
4. EYSENCK, H. J. Primary social attitudes: a comparison of attitude patterns in England, Germany, and Sweden. *J. abnorm. soc. Psychol.*, 1953, **48**, 563–568.
5. EYSENCK, H. J. *The psychology of politics*. London: Routledge & Kegan Paul, 1954.
6. FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.
7. KEEHN, J. D. An examination of the two-factor theory of social attitudes in a Near Eastern culture. *J. soc. Psychol.*, 1955, **42**, 13–20.
8. PROTHRO, E. T. Generalized ethnic attitudes in the Arab Near East. *Sociol. soc. Res.*, 1953, **37**, 375–379.
9. PROTHRO, E. T., & MELIKIAN, L. H. The California Public Opinion Scale in an authoritarian culture. *Publ. Opin. Quart.*, 1953, **17**, 353–362.
10. SANFORD, N. Recent developments in connection with the investigation of the authoritarian personality. *Sociol. Rev.*, 1954, **2**, 11–33.
11. SHILS, E. A. Authoritarianism: "right" and "left." In R. Christie & Marie Jahoda (Eds). *Studies in the scope and method of "The Authoritarian Personality."* Glencoe, Ill.: Free Press, 1954. Pp. 24–49.
12. WEBSTER, H., SANFORD, N., & FREEDMAN, M. A new instrument for studying authoritarianism in personality. *J. Psychol.*, 1955, **40**, 73–84.

Received September 27, 1955.

RORSCHACH AND TAT INDICES OF HOMOSEXUALITY IN OVERT HOMOSEXUALS, NEUROTICS, AND NORMAL MALES[1]

ANTHONY DAVIDS

Brown University and Emma Pendleton Bradley Home

MARK JOELSON AND CHARLES McARTHUR

Harvard University

THERE has always been much concern regarding the nature, cause, and prevalence of homosexual behavior. Although the extent of this behavior in contemporary America can be estimated only crudely, reports by social scientists suggest that homosexual activity of some kind, prepubertal, adolescent, or adult, is far more prevalent than is generally realized (3, 5).

Although the diagnosis of homosexuality is not difficult to make for those individuals who do not attempt to hide their abnormality, there are many persons who recognize and perhaps act on desires which are more than latent, yet who deny them. The attitudes of society and demands of social living cause these individuals to retreat alone into the shadows with their problem. Thus, the diagnosis of homosexuality by ordinary direct methods of assessment is often difficult or impossible. The development of projective techniques, however, opened up new avenues for diagnosis and analysis of psychological dynamics. The present investigation, using the Rorschach test and the Thematic Apperception Test (TAT), is an attempt to contribute to further understanding and clarification of the problem of homosexuality in young adult males.

More specifically, this study, employing a group of overt homosexuals, a group of patients suffering from other neurotic difficulties, and a control group of normal subjects (*S*s), will:

1. Compare the Rorschach protocols of these groups in regard to homosexual signs derived from the psychological literature.

2. Compare the TAT protocols of these groups in regard to homosexual signs derived from the psychological literature.

3. Compare and correlate the results on both tests within each of the groups.

4. Suggest for further research any new signs uncovered in this study that show a significant difference between the groups.

[1] This study was facilitated by the Laboratory of Social Relations and the Department of Hygiene at Harvard University, and by research grant M-700 from the National Institute of Mental Health, Public Health Service.

METHOD

Subjects

Homosexual group. The overtly homosexual group was composed of 20 men who, at the time of testing, were students attending large Eastern universities. They were all concerned about their problem, wanted to talk to the psychologist about it, and were aware that the psychologist knew they were overt homosexuals. These men were all tested by the same psychologist in his office in a university clinic setting. Their intelligence ranged from high average to superior, and all came from average or better socioeconomic backgrounds.

Neurotic group. The neurotic group was composed of 20 male students who were attending a large Eastern university and were being seen for psychological diagnosis and psychiatric treatment at the university's hygiene department. These *S*s were suffering from neurotic emotional disturbances. If diagnostic labels were to be used, the group would consist mainly of anxiety states, obsessive-compulsives, hysterics, and depressives. All of these patients were of high average or superior intelligence, and came from average or better socioeconomic backgrounds. The examining psychologist, who also tested the homosexuals, affirmed that there were no known homosexuals in this group. In the present study, these emotionally disturbed *S*s, who were also tested under conditions conducive to frankness, constitute a control group. That is, this group was introduced to ensure that differences found between the homosexuals and the normal control *S*s, described below, were not a function of a generally disturbed state but were reflections of specifically homosexual disturbance.

Normal group. The normal *S*s were 20 male students attending a large Eastern university and, at the time, participating in an intensive investigation of personality being conducted by a team of trained investigators. These *S*s were selected by an experienced clinical psychologist who tried to secure a sample of the college population containing a diversity of interests, abilities, and personalities. The men were above average in intelligence and came from average or better socioeconomic backgrounds. They had all volunteered to serve as *S*s for the research project, participated in it over a two-year period, were guaranteed anonymity, and were told that the data would be used solely for purposes of sci-

entific research. Therefore, it is reasonable to expect their responses to be honest and valid.

From these descriptions, it is evident that the three groups were essentially comparable in respect to most significant variables. For purposes of the present study, it is important to point out that there were no significant differences between the groups in regard to college major. Each group appeared to contain a fairly representative sampling of the various fields of concentration, with few *Ss* in each group majoring in subjects indicative of strong aesthetic interests. It can be stated with considerable certainty that the groups did not differ significantly in regard to literary or aesthetic interests. The one major variable that differentiated these groups was the overtness of homosexual practices in one, whereas the other two, as far as it could be ascertained, contained no overt homosexuals.

The Rorschach Test and Thematic Apperception Test

Administration. Standard instructions and procedures for administration of the TAT (9) and the Ror-

TABLE 1

RORSCHACH SIGNS OF HOMOSEXUALITY*

Sign	Card	Location	Content	Examples
1	I	Whole or part of whole	Mask or human or animal face	"Theatrical mask" "Fox, leering"
2	I	Lower center detail	Male or muscular female torso	"Japanese man" "Woman athlete"
3	II	Lower center detail	Crab or crablike animal	"Deep-sea crab" "Cooked lobster"
4	III	Whole or part of whole	Humans; with sex confused	"Woman above, man below," "Man, but with breasts"
5	III	Whole or part of whole	Humans; with sex uncertain	"Two sexless people" "Could be either"
6	III	Whole or part of whole	Animal or animal like (dehumanized)	"Featherless chickens" "Ostrich-like natives"
7	IV	Whole or part of whole	Human or animal; contorted, monstrous, or threatening	"Man looking back through legs," "Horrid beast"
8	V	Whole, part of whole, or center detail	Human or humanized animal	"Woman dressed as bat" "Bugs Bunny"
9	VI	Center or top detail	Object; with implication of cleavage	"Ship with foaming wake" "Oil well, drilling"
10	VII	Whole, part of whole, or top detail	Human; female with derogatory specification	"Two old hags" "Women—yakety-yak"
11	VIII	Lateral detail	Animal; several incongruous ones or one with incongruous parts	"Rat or a lion or something," "Bear with a rat's head"
12	IX	Upper lateral detail	Human; dehumanized	"Witch-like creature" "Monster with guns for arms"
13	X	Top center detail	Animal; attacking or fighting over central object	"Beavers gnawing a tree," "Ants arguing over this"
14	X	All or upper half of pink plus center blue	Human; with blue as oral specification	"Ectoplasm coming from mouth," "Irishman with growth on lips"
15	Any card, I–X	Any location	Human or animal	"Decayed teeth" "Open mouth showing tonsils"
16	Any card, I–X	Any location	Human or animal anal detail or specification	"Woman with huge ass" "A bloody rectum"
17	Any card, I–X	Any location	Human or animal described as "back to back"	"Two Negroes with their backs pressed together" "Two sheep back to back"
18	Any card, I–X	Any location	Human object or architecture with religious specification	"A priest's cross" "Two men praying"
19	Any card, I–X	Any location	Male or female genitalia	"Part of a man; his nuts," "A woman's private parts"
20	Any card, I–X	Any location	Feminine clothing	"A woman's fur jacket" "Panties"

* Adapted from Wheeler (13, pp. 104–105).

schach (6) were followed. The complete Rorschach test (10 cards) was administered to all *Ss* in the three groups. However, there was some variability as to which TAT cards were administered to different *Ss* as well as differences in the manner of recording the stories. The TAT stories of the homosexual and neurotic groups were recorded, using a form of shorthand and later typed in detail in the attempt to secure verbatim protocols. The stories of the normal group were electrically recorded on Dictaphone belts and were transcribed verbatim. As for cards administered, there was considerable inconsistency in the homosexual group. The TAT records for this group ranged from one card (the test being abandoned when the *S* became disturbed) to 15 cards. It was possible, however, to form subsamples, within the homosexual group, of those who had all responded to the same stimulus cards. Fifteen of the 20 homosexuals responded to Cards 4, 6, 7, 10, and 18 of the male series. Also, five of the homosexuals were administered Cards 1, 3, 4, 6, 10, 11, 13, 14, and 18, which permitted further comparisons within this more limited sample. Thus, a few of the "signs" derived from the TAT literature could not be adequately tested because the required card had not been administered to the homosexual group. The neurotic group responded to 10 TAT cards: 1, 3, 4, 6, 7, 8, 13, 14, 17, and 18; the normal group composed stories for the entire set of 20 cards in the male series of the TAT. In spite of this lack of consistency of cards administered to the *Ss*, the available TAT data proved quite satisfactory for purposes of the present investigation.

The Rorschach signs and their scoring. In considering which Rorschach indices of homosexuality, taken from the psychological literature, offered the most promise, it was decided that the list compiled by Wheeler (13), drawn in part from three earlier studies (2, 4, 7), represented the most complete and specific scheme available. Although the Wheeler scheme has been utilized elsewhere, it had never been put to a crucial test with a group composed entirely of overt homosexuals. Detailed description of these signs, their derivations, and rationale, is presented in Wheeler's report (13). An abbreviated description of the 20 signs is presented in Table 1. It is apparent that the majority of these signs involve a minimum of ambiguity in their scoring. The data were scored "blind," so that the experimenter did not know if he was scoring the protocol of a homosexual or a control *S*. The scorer went through each protocol scoring the presence or absence of each sign and recording the total number of homosexual signs in the entire protocol.

The TAT signs and their scoring. There is a dearth of organized schemes containing TAT indices of homosexuality. Tejessy (11), however, has presented a list of empirical generalizations concerning "typically homosexual" responses as suggested by scattered statements in the TAT literature. In a later paper, Lindzey, Tejessy, Davids, and Heinemann (8) modified and lengthened the list of empirical generalizations and quoted additional authorities for some of the indices previously reported. Detailed descriptions of the 10 homosexual signs, presented briefly in Table 2, with a discussion of their source and rationale, is contained an these two manuscripts (8, 11). The present investigators, after a careful survey of the literature, were unable to find any published reports of the experimental test of the presence of these signs in protocols of overt homosexuals. The TAT records for the three groups were scored "blind," the scorer recording the presence or absence of each sign and the total number of homosexual signs in the protocol.

Reliability of scoring. An independent investigator scored the Rorschach signs in 10 of the homosexual protocols and 10 of the normal group protocols. The interrater agreement of scoring was 83 per cent; the product-moment correlation between the number of signs scored by each rater was .92. Another independent judge scored the TAT signs in 10 of the homosexual protocols and 10 of the normal group protocols. The interrater agreement of scoring was 96 per cent, and the product-moment correlation between the number of signs scored by each rater was .85.

Hypotheses

Hypothesis I. The homosexual group gives a significantly greater mean number of proposed homosexual Rorschach signs than does the neurotic group or the normal group.

Hypothesis II. The homosexual group gives a significantly greater mean number of proposed homosexual TAT signs than does the neurotic group or the normal group.

Hypothesis III. Within each group there is a significant positive correlation between Rorschach and TAT signs of homosexuality.

RESULTS

Rorschach Test

In support of Hypothesis I, the homosexual group gave a significantly greater mean number of Rorschach signs proposed as indices of homosexuality than did either of the control groups. As shown in Table 3, the homosexual group gave a mean of 5.2 signs, while the normal control group gave a mean of 2.8, and the neurotic group a mean of 2.3 signs. For both comparisons, the difference

TABLE 2

TAT SIGNS OF HOMOSEXUALITY

Sign	Card	Content
1	9	Card makes subject uncomfortable
2	10	Subject thinks that scene is a sham, that there is really no reciprocal affection
3	12	Stories in which the young man has been hypnotized by the older man
4	18	Stories in which the hero has been attacked from the rear or pulled to the rear
5	Any card	A high degree of feminine identification
6	Any card	A recurrent shift of identification
7	Any card	Misrecognitions of sex of TAT figures
8	Any card	Derogatory attitude toward marriage and the opposite sex, including stories of marital discord
9	Any card	Stories in which men kill women
10	Any card	Sexual, genital references

TABLE 3

RORSCHACH SIGNS OF HOMOSEXUALITY AND PRODUCTIVITY OF RESPONSES

	Homosexual Group (N = 20)		Neurotic Group (N = 20)		Normal Group (N = 20)		Homosexual vs. Neurotic		Homosexual vs. Normal		Neurotic vs. Normal	
	Mean	Variance	Mean	Variance	Mean	Variance	F	t	F	t	F	t
Number of signs	5.2	6.7	2.3	5.4	2.8	1.9	1.24	3.76**	3.54**	3.66**	2.84*	.79
Number of responses	67.9	1860.8	45.0	798.2	50.6	385.3	2.33**	1.99*	4.83**	1.63	2.07	.73
% signs per response	9.3	22.5	5.3	30.6	6.3	17.6	1.36	2.45*	1.28	2.13*	1.74	.65

* Significant at the .05 level.
** Significant at the .01 level.

between the homosexuals and the nonhomosexual *S*s is significant well beyond the .01 level. However, the means of the two nonhomosexual groups do not differ significantly.

In order to determine whether the significant difference in signs might have been caused, in part, by a difference in total responses, the difference between the mean number of responses given by each group was measured. The results presented in Table 3 show that the mean number of responses in the homosexual group is significantly higher than the mean in the neurotic group, but not significantly higher than the mean in the normal group when evaluated according to a two-tailed test of significance. A test of the difference between the neurotic and normal groups, as shown in Table 3, was not significant.

Since the homosexual group gave more signs and also produced more responses than the control groups, the mean percentage of signs given in proportion to total number of responses was computed for each of the groups. It was found, as shown in Table 3, that a significantly greater percentage of the homosexual group responses were homosexual signs, and that the mean percentage of homosexual signs did not differ significantly between the control groups. Further evidence that sign production is not necessarily related to total productivity is furnished by the fact that the mean number of homosexual signs given by the three homosexual *S*s who produced the most total responses was not as high as the mean number of signs for the homosexual group as a whole.

Thus, Wheeler's scheme of 20 signs seemed to be successful in differentiating overt homosexuals from neurotics and normal males, while not differentiating between the latter two groups. In addition, it was found that the overt homosexuals produced an average of 23 more responses to the Rorschach test than did the hygiene group, and 17 more responses than did the normal *S*s. This finding is in agreement with Aronson's (1) suggestion that creative productivity on the Rorschach may be a homosexual variable. Since, with productivity controlled, the mean percentage of signs for the homosexual group was significantly higher than the means for the other two groups, it is concluded that the high response productivity of the homosexuals was not a vital factor in their production of more proposed signs of homosexuality.

Validity of individual signs. Although the 20 signs, as a scheme, differentiated between the homosexual and nonhomosexual groups, only 4 of the individual signs discriminated significantly between the homosexuals and the normal *S*s. These individual signs are *Sign 7* (t = 1.70), *Sign 8* (t = 2.17), *Sign 10* (t = 1.65), and *Sign 19* (t = 1.80). None of these signs differentiated significantly between the two nonhomosexual groups.

The validity of *Sign 7* (Blot IV seen as contorted, monstrous, or threatening) is consistent with the clinical interpretation of Card IV as a stimulus that frequently evokes signs of castration anxiety. It is not readily apparent from the rationale offered by Wheeler why *Sign 8* (humanization of animal) should be so successful, but both Wheeler (13) and

Reitzell (10) identified it as an effective sign. *Sign 10* (derogatory attitude toward females) and *Sign 19* (preoccupation with genitalia) have an obvious relation to homosexual dynamics.

It should be emphasized that these signs are listed because they show a *significant degree of differentiation*, not because of their *frequency* of occurrence in the homosexual protocols. That is, the frequency of occurrence of a sign was not always indicative of its ability to differentiate.

Most of the remaining signs proved useful, to some degree, although many of them did not occur frequently enough to be validated. Only one sign can be seriously questioned. *Sign 1* (mask or face on Card I) was given 11 times by the homosexual group and 11 times by the normal group. The frequency of occurrence of this sign, in normal as well as homosexual *S*s' protocols, suggests that intellectualization of Card I into a face may not be a valid discriminator, and the usefulness of this sign awaits further experimental evidence.

Validity of individual cards. Four of the 10 Rorschach cards differentiated between the homosexual group and the normal group, in the predicted direction, beyond the .05 level of significance. These are Cards II ($t = 1.97$), IV ($t = 2.05$), V ($t = 1.80$), and VII ($t = 1.70$). None of these cards differentiated significantly between the two nonhomosexual groups. The results obtained with Cards IV and V were due largely to the success of *Signs 7 and 8*, which accompany these cards. The present findings with Card IV support the clinician's faith in this card as securing responses indicative of confusion and unresolved conflicts in regard to relations with males in general and "father figures" in particular. Card II elicited genital, anal, and religious responses, while Card VII evoked responses derogatory to women, and descriptions of animals or humans as being "back to back."

Signs suggested for further confirmation. It was noted that the homosexual *S*s gave many responses involving the *rear view* of human beings or animals. For example, responses like "figure seated with its back to you," "back of two legs," or "Negro can-can," were given quite frequently by the homosexuals. Comparison of the mean frequency of rear-view responses given by the homosexual group and the normal control group resulted in the highly significant t of 2.53. This *rear-view* sign is probably a more defended form of Wheeler's *anal detail* sign. The latter sign did not occur very frequently in the present study. At any rate, the finding of a significant difference between the homosexual and normal *S*s suggests that the rear-view sign may prove to be a valid index of homosexuality. This new sign is offered for further investigation and confirmation.

Discussion of Rorschach findings. Although content analysis seemed useful here, certain deficiencies of this type of analysis became apparent to the investigators. Much of the distinctiveness of the homosexual protocols could not be measured by the sign method. The homosexual reactions to the blots, as reflected by personal remarks and the use of particular words recorded in the data, seemed to display a great sensitivity of the *S* toward the blots and a need to relate each blot to himself. Remarks such as "I don't like that one," or "he looks mean," were frequent for the homosexual *S*s, while the control *S*s gave responses that were more detached. This sensitivity and greater ego-involvement toward the Rorschach test is, in part, reflected by the greater productivity of the homosexuals. Also, the homosexuals displayed a greater "shock" reaction to color on the Rorschach than did the control *S*s. These and other nuances, which prove most useful to the clinician, are overlooked by systems, such as the present one, of sign scoring.

As a whole, however, the signs were successful in differentiating between the homosexual group and the control groups, but the dangers of applying the scheme blindly to an individual protocol were obvious. One member of the normal group gave seven signs, while another gave six, both being above the mean number given by the homosexuals. Examination of the complete case records of these two control *S*s revealed no evidence of homosexual tendencies. On the other hand, one homosexual *S* gave only two sign responses, which is below the mean for the normal *S*s. Thus, there is considerable danger of making a grave mistake in diagnosis if the signs are applied as criteria without supplementary information.

TABLE 4
VALIDITY OF TAT SIGNS OF HOMOSEXUALITY

Cards 4, 6, 7, 10, 18	Homosexual Group ($N = 15$)	Normal Group ($N = 20$)	F	t
Mean number of signs	2.73	.85		3.71**
Variance	3.21	.87	3.70**	
Cards 4, 6, 7, 18	**Homosexual Group ($N = 15$)**	**Neurotic Group ($N = 20$)**	**F**	**t**
Mean number of signs	1.67	.63		4.08**
Variance	.95	.36	2.64*	
Cards 1, 3, 4, 6, 10, 11, 13, 14, 16, 18	**Homosexual Group ($N = 5$)**	**Normal Group ($N = 20$)**	**F**	**t**
Mean number of signs	3.80	2.40		1.87*
Variance	2.20	2.25	1.02	

* Significant at the .05 level for a one-tailed test.
** Significant at the .01 level for a one-tailed test.

Thematic Apperception Test

The lack of consistency in administering TAT cards to the various *S*s has been mentioned. However, several comparisons were made possible by forming subsamples from the various groups, the size of these subsamples and the number of cards employed varying in each case in order to make the most efficient use of the available data.

A comparison involving the greatest possible number of *S*s was made between 15 homosexual *S*s and 20 normal *S*s on the basis of responses given to cards 4, 6, 7, 10, and 18. With this equitable five-card comparison, the homosexual group mean of 2.7 signs was significantly higher, as shown in Table 4, than the normal control group mean of .9 signs.

The best comparison possible employing the neurotic group was made on the basis of responses given to Cards 4, 6, 7, and 18 by the same 15 homosexual *S*s and the 20 neurotic *S*s. As shown in Table 4, for these four cards, the homosexual group mean of 1.7 was significantly greater than the neurotic group mean of .6 signs.

These two comparisons, although based on responses made to a small number of cards, do employ cards that were selected because they seemed most likely to elicit material relevant to homosexuality. The special relevance of Card 10 is evidenced by the considerable reduction in the homosexual group mean in the second comparison. In any event, these two comparisons clearly confirm Hypothesis II.

One additional comparison was made in an effort to take into consideration responses given to a greater number of cards. Five of the homosexual *S*s had composed stories for Cards 1, 3, 4, 6, 10, 11, 13, 14, 16, and 18, and these responses were compared with the stories given by the 20 normal control *S*s in response to these same 10 cards. A t of 1.87 resulted, which is statistically significant but considerably lower than the other t values presented in Table 4. On closer examination, two factors seemed to account for the reduced t. The inclusion of Card 13 enabled the normal *S*s to give numerous sexual references, while the inclusion of Card 3 led to increased misrecognition of sex in this normal group, a distortive factor which will be discussed later. These sexual references and misrecognitions, which were encouraged by the nature of the stimulus pictures, although also increasing the number of signs produced by the homosexual *S*s, decreased the significance of the proportional difference in signs between the two groups.

Validity of individual signs. Since comparisons were based on only a few cards, and since the TAT signs did not occur as frequently as did those in the Rorschach protocols, it could not be expected that many of the TAT signs would individually be capable of differentiating between the groups. However, two signs (*Signs 5* and *10*) did discriminate significantly. The results with these two signs are based on the five-card comparison (Cards 4, 6, 7, 10, and 18) between the 15 *S*s in the homosexual group and the 20 *S*s in the normal control group.

The most discriminatory sign was *Sign 10* (sexual, genital references). The homosexuals gave more sexual references ($t = 3.46$) than did the normal *S*s. Here, a particularly note-

worthy finding was that among the sexual signs scored for the homosexual *S*s on these five cards, there were seven explicit homosexual stories. Moreover, two additional stories of homosexual relations were told in response to other cards. These nine stories were told by six of the homosexual *S*s. *No* such stories were found in any of the data available on either of the two nonhomosexual groups. These homosexual stories revealed a wealth of information about the dynamics of the storyteller, but their diagnostic value is undoubtedly of limited generality, since such stories are likely to be told only by admitted overt homosexuals who are aware that the examiner knows the nature of their problem. However, a t test between the number of sexual responses given by the two groups, omitting the seven stories of overt homosexual activities, still resulted in a significant t of 3.17.

The other discriminating sign was *Sign 5* (a high degree of feminine identification). The t in this case was 1.71, significant at the .05 level for a one-tailed test. The success of this sign is particularly meaningful for two reasons. Since identification is often difficult to distinguish from an objective attempt by the *S* to describe the pictured character fully and realistically, identification was only scored when feminine identification was quite pronounced and the storyteller showed an unusual ability to create a picture of feminine yearnings and feelings. Furthermore, of the five cards used for comparison, the three that have female characters in the situation also contain equally prominent male characters, so that sexual identification was entirely a function of the *S*. This is in contrast to a card like Card 2, in which a woman is centered in the picture, providing an obvious choice for identification. At any rate, the present finding of a greater need and ability of the homosexual *S*s to identify with females is in agreement with psychological theory on homosexuality.

Sign 7 (misrecognition of sex) presents special difficulties, notably in regard to its association with Card 3. According to the TAT manual (9), the character portrayed in Card 3 is labeled as a boy; thus, any reference to this character as a female was scored as a misrecognition of sex. However, the possible ambiguity of the stimulus figure in this picture was sharply pointed out by the fact that 9 of the 20 *S*s in the normal control group described the character as a girl or woman. In addition, 10 of the 20 neurotic *S*s described it as a woman, while 8 of the 11 homosexuals who responded to this card called it a woman. The incorrectness of scoring this as a sign of homosexuality is clearly illustrated by a comparison of the three proportions, which results in no significant differences. These findings suggest quite definitely that the figure in Card 3 may best be considered an ambiguous figure. Few misrecognitions were made by any of the *S*s in response to other cards.

Therefore, the TAT signs, as a scheme, seemed to be characteristic of homosexuality, although only two of them individually discriminated significantly. These two, sexual references and feminine identifications, were found especially useful when applied to Cards 4, 6, 7, 10, and 18, which, unlike some of the others, do not confront the *S* with a stimulus situation necessarily requiring a sexual (genital) plot or identification with a female. The validity of the other signs was found to be either limited or unsatisfactory. However, the distinctive quality of the homosexual protocols suggested that other differentiating signs were there to be found.

Validity of individual cards. Of the five cards that, because of their theoretical relevance, were administered uniformly to the majority of the homosexuals, four attracted significantly ($p = .05$) more homosexual signs from the homosexual group than from the normal control group. The best card proved to be Card 10 ($t = 2.49$), which elicited sexual references, feminine identifications, and stories of marital discord and pretense from many of the homosexual *S*s, while most of the normal *S*s composed "happy marriage" stories, and few of them identified with the woman in the picture. Card 7 was also successful in differentiating between the groups ($t = 2.27$), although much of this success came from its ability to elicit stories of overt homosexual relations from the homosexual *S*s. Cards 4 and 6 also differentiated significantly ($t = 1.77$ and 1.90 respectively), again mainly because of the sexual references, feminine identifications, and derogatory attitudes toward women and marriage. Card 18 resulted in differences in the predicted direction, but did not dis-

criminate significantly ($t = 1.32$). It is possible, of course, that some cards in the female series (e.g., 18GF) might be worth experimenting with in future studies of male homosexuality.

Signs suggested for further confirmation. Five new signs were suggested for further validation, three of them significantly differentiating homosexuals from normals, and the other two not occurring as frequently but worthy of mention for their possible diagnostic value.

The first sign suggested for further research applies to stories containing a *strong unresolved attachment to a father or father figure.* Scored in stories composed for Card 8, this sign resulted in the highly significant t of 2.54. It was only scored when the attachment to the older man was the focal point, not merely incidental to other facets of the story. Remarks such as, "He couldn't bear to leave his father," and "The older man and the younger man had a strong relationship, because the old man took a great deal of interest in him," were frequent in the homosexual protocols. This suggested sign has both a theoretical and an empirical basis, the observation of homosexual relationships between a young man and an old man being theoretically referable to the unresolved Oedipal conflict.

Another sign, associated with Card 6, applied to stories involving a *strong unresolved attachment to the mother.* Comparison of the homosexuals and the normal *Ss* resulted in a significant t of 1.96. Here again, the sign was only scored when the mother-son relationship was described as being "a silver cord attachment," or from statements like, "They had always been too close, if anything." Moreover, one explicit mother-son incest story was produced by a homosexual *S*. Although the normal *Ss* told of mother-son relationships in response to Card 6, the focus was usually a financial problem, a "guilt" theme the boy's leaving his mother to marry, or a sense of duty or feeling of unworthiness in the son's relationship to the mother. Whenever a story implied an unnatural love toward the mother, it was found to be produced by a homosexual *S*. The success of this sign is understandable, theoretically, in terms of an unsatisfactory resolution of the Oedipal love for the mother.

Although the homosexuals did not often produce stories in which a clearly derogatory attitude toward women was revealed, use of certain *sexual derogatory terms in relation to women* gave specific clues to their feelings toward female characters in their stories. Words like "tramp," "babe," "temptress," "vamp," "floozy," "wench," and so forth were far more frequent in the homosexual protocols than in the protocols of the normal *Ss*. A comparison of the mean number of such terms used by the homosexuals and the normal *Ss* resulted in a significant t of 2.31. It is noteworthy that in many of the stories employing these words, the homosexuals described another female character in a favorable light, so that the derogatory attitude was not evident in the story as a whole, but was betrayed only by the occasional word.

Two other possible signs, which did not occur with sufficient frequency to be validated but were notably absent from the normal *Ss*' stories, seem worthy of further investigation. The homosexuals gave several stories in which a *heterosexual relationship was followed by tragedy,* usually the violent and sudden death of one member of the couple. Also, in response to Card 18, several homosexuals told stories that *describe the plight of the young man allegorically or symbolically,* defining the situation as a "symbol of revolt against society" or the "torture of the damned." This allegorical description, which is in contrast to the realistic stories told by the normal *Ss*, is possibly an intellectualization of the *Ss*' homosexual conflict.

Discussion of TAT findings. As in the Rorschach, the examiner could often feel an immeasurable distinctiveness in the homosexual TAT data. This feeling was unbiased since the stories were first scored blind, with the scorer able only to guess to which group a protocol belonged. There was, again, in the homosexual protocols, a deeper sensitivity than was displayed by most of the normal *Ss*, a fuller projection into the story, including the ability to empathize with the characters. This sensitivity was accompanied by elaborate descriptions of the emotional relationships between various characters in the story. Also, there was often a sense of fatalism, of tragic events to come which could not be evaded by the characters. These qualities, which are difficult to describe objectively and to quantify, are meaningful to the clinician but are difficult to record by the present sign scheme.

Again, the uses and limitations of sign

analysis are evident. The success of the scheme in differentiating between groups is offset by the dangers involved in judging the individual case by the number of homosexual signs in the protocol. One normal *S* gave seven TAT signs of homosexuality, considerably above the mean number for the homosexuals, yet his complete file was found to be free of evidence of homosexuality. As with the Rorschach, TAT signs should be used only in conjunction with supplementary evidence gained from the case history and other psychodiagnostic procedures. They definitely cannot be expected to carry the burden of complete diagnosis of the individual case.

Joint Use of the Rorschach and Thematic Apperception Test

Hypothesis III predicted a significant positive correlation within groups between the number of Rorschach and TAT signs of homosexuality given by each *S*. It would be expected that if the Rorschach and TAT signs were each capable of detecting this abnormality, they would agree in their measurement of the relative prominence of the variable in each *S*.

Product-moment correlations (r) were computed within each group to determine the degree of association between results of the two tests. With the neurotic group, the resulting r of $-.09$ does not permit rejecting the null hypothesis of no association; and with the normal control group, the correlation of .05 fell far short of significance.

Since the degree of correlation of the two tests for the homosexual group is of crucial interest in this study, the TAT results were ordered into two schemes for correlation with the Rorschach signs. First, the complete sign totals on both tests for all 20 *S*s were correlated, although the TAT signs of many of these *S*s were based on a varying number of cards, ranging from one to fifteen. It could be expected that the disparity in the completeness of the TAT data on each *S* might prove detrimental to obtaining the predicted positive relationship. However, in spite of this inequity, a significant r of .52 resulted from this comparison.

Then a more equitable correlation was attempted within the homosexual group. The TAT sign totals of the 15 *S*s who had produced stories for the same five cards were correlated with the sign totals of these *S*s on the Rorschach test, resulting in a coefficient of .51. Thus, both correlations for the homosexual group proved positive and significant.

TABLE 5

CORRELATION BETWEEN HOMOSEXUAL SIGNS ON THE RORSCHACH AND TAT

Group	*N*	*r*
Neurotic	20	−.09
Normal	20	+.05
Homosexual	20	+.52**
Homosexual	15	+.51*

* Significant at the .05 level for a one-tailed test.
** Significant at the .01 level for a one-tailed test.

Although Hypothesis III was confirmed only in the case of the homosexual *S*s and unsupported by the records of the control *S*s, these results, shown in Table 5, provide a clear insight into the success and limitations of this investigation. Apparently the two tests did what they were expected to do. They discriminated a group of homosexuals from two groups of nonhomosexual males. In addition, they measured the relative prominence of the homosexual characteristic in a similar fashion for the homosexual *S*s. However, to expect the tests to correlate significantly in detecting signs of latent homosexuality in nonhomosexual *S*s is not only drawing too fine a line for differentiation, but also entails the assumption that each sign is a relatively foolproof detector of the homosexual variable. The results of this experiment certainly demonstrate the fallacy of considering any sign a definite detector, as well as point out the varying validity of the signs. In the case of the homosexuals, the possible lack of validity of a given sign is usually offset by the large total number of signs, which is somewhat more valid and less susceptible to error.

DISCUSSION AND CONCLUSIONS

While these two schemes appear useful in characterizing homosexuality, there are, nevertheless, certain limitations that must be kept in mind, both in accepting the validity of these indices and in deriving generalizations from them. The most glaring and serious limitation is that these results do not permit the conclusion that a large number of these signs in the Rorschach or TAT is necessarily an indication of homosexuality. The case of

the normal *S* with several signs in his protocol, as well as that of the homosexual *S* with very few signs, definitely occurred. This possibility is to be expected, since several of the signs, in either scheme, cannot be considered unusual responses, and some could be interpreted in terms of other personality variables. For example, in one record "anal" imagery may be symptomatic of homosexuality, while in another the same imagery may be indicative of an obsessional theme. Even feminine identification, a useful sign in the TAT and the rationale for a Rorschach sign, when based on identification with the mother, is not infrequent in American society and is reflected in the TAT stories of many heterosexual males (12).

It is possible that some signs proved useful in this study only because the homosexual *S*s were highly intelligent. Perhaps, for example, a group of homosexuals of different intellectual capacity would not humanize animals on the Rorschach test. Since the present design does not permit the rejection of such possibilities, it seems wise to restrict the import of these findings to men of college age, of high average socioeconomic background, and of high intelligence. However, as the homosexual variable was still the most basic, it seems plausible that these results and considerations may hold true for homosexuality in general.

The use of a group of overt homosexuals permitted a critical test of the validity of the proposed signs. But the use of this group of *S*s also presents a difficult problem in the interpretation of the results. Since these *S*s all came to be tested because they wanted to talk about their problem and all were aware that the examiner knew their reason for coming, it is to be expected that their test responses would be frank and unsuppressed. Therefore, much of the knowledge gained from studying this homosexual group may be of little avail in investigating the protocols of other individuals who have no desire to be frank and who may feel themselves at odds with the psychologist. In other words, it is difficult to say how much of what has been learned here will prove useful in detecting a latent homosexual conflict or in identifying a homosexual who is deliberately trying to prevent discovery of his problem.

In reviewing both the historical background of the problem and the results of this study, it has become obvious that Rorschach research on homosexuality has been more complete and effective than what has been done with the TAT. The Wheeler scheme, in comparison with the TAT list of signs, is more complete, more specific, and seems to differentiate better between the groups.

Limitations of the TAT sign approach were made clearly evident by several examples in the protocols. A bizarre story told by a homosexual in response to Card 13 about a boy who shares his girl with his roomate is scored as a sexual reference along with other more conventional sexual stories told by control *S*s. Full use and interpretation of such meaningful material cannot be handled by a system of sign scoring, especially one so crude as that used in this experiment. However, the distinctive quality of the homosexual TAT responses suggests that additional signs of a definite nature can be postulated and may prove useful. This study has suggested five such signs. If, in a future study, a greater number of TAT cards is administered to a homosexual group, more such relevant material would probably be uncovered.

The success of various signs in both projective tests contains several implications for theory on the topic of homosexuality. A preoccupation and lack of satisfaction with sex is brought out by the many genital responses given on the Rorschach by the homosexuals as well as by their inclusion of sexual themes and derogatory sexual terms on the TAT in response to stimuli which are not specifically structured to elicit such material. The derogatory attitude of these homosexuals toward women in general is evidenced both on the TAT and the Rorschach. Paradoxically, this depreciation of women seems to be often accompanied by a feeling of identity with them, manifested not only on the TAT, but on the Rorschach as well.

Oedipal implications are involved in the stories of unnatural love and overclose ties to the mother. An interesting note here is that in most of the stories told by homosexuals of this abnormal closeness to the mother, there was usually a deprecating tone in reference to the attachment, as if the *S* were admitting this tie grudgingly and as if he were trying desperately to break away from it. This is man-

ifested in references such as "the umbilical cord was never broken." In contrast to the depreciation of the overclose mother tie, bonds to another man were usually idealized. The older man was "very kind," or "really interested in him," and there was contentment for both parties in the male relationship. Thus Oedipal implications concerning the homosexual *S*'s identification with the mother are evidenced by these stories of "love for an older man." The importance of parental figures in shaping later attitudes toward and identification with the sexes was brought out again and again.

The confusion and conflict in the homosexual role is indicated by the success of several of the signs. The frequent religious references in response to the Rorschach inkblots suggest the heavy burden of guilt and self-questioning that the homosexual must bear. The allegorical references on the TAT are probably also an expression of the homosexual's inner conflict. The humanization of animals on the Rorschach test may also suggest a confusion of role and sexual identity, or even of body image, as a source of anxiety.

One further point seems worthy of mention. It was observed that the responses of the homosexuals on both tests ranged from the most romantic and unrealistic, such as symbols of royalty on the Rorschach and tales of the "tragic hero" and stories of "another world," to the most lurid and earthy Rorschach responses like "bleeding rectum" or "vaginal disease" and TAT stories of incest and suicide. This approach to the projective stimuli may represent a manifestation of the homosexual conflict between high sensitivity and craving for beauty and all the disgust and disfavor which society associates with the homosexual's problem. Both extremes imply a lack of realism in thinking.

This discussion points to the wide possibilities of both a "sign" analysis and a more global clinical appraisal of projective protocols. The Rorschach research has progressed further than that with the TAT, but additional investigation and validation of indices on both tests is needed. As the list of valid signs of homosexuality increases and unsatisfactory ones are eliminated, this approach to diagnosis and understanding of the homosexual problem will become less crude, and may well have more implications for the individual case. At present, however, this content "sign" analysis should only be used to supplement the more structural and inferential appraisal of projective protocols in conjunction with any other diagnostic material available.

SUMMARY

Rorschach and TAT protocols of a group of overt homosexuals, a group of neurotics, and a group of normal males were scored for frequency of occurrence of proposed signs of homosexuality. As hypothesized, it was found that the homosexual group gave a significantly greater mean number of the Rorschach signs than did either nonhomosexual group. Four individual signs, and four separate Rorschach cards were found to differentiate significantly between the homosexual and the normal group. One new Rorschach sign uncovered in this study was offered for further confirmation. As predicted, it was also found that the homosexuals gave a significantly greater mean number of the proposed TAT signs than did either nonhomosexual group. Two individual signs were significantly discriminatory, and five separate TAT cards were recommended for administration in future studies of homosexuality. Five new signs were suggested, along with their theoretical implications, as aids to formulation of a more discriminating scheme.

Within the homosexual group, correlation between the number of Rorschach signs and number of TAT signs produced by each *S* proved significant, serving as a check on the validity of both schemes and indicating the consistency of these two diverse measures of homosexuality. However, similar correlations within each of the nonhomosexual groups did not approach a level of statistical significance, indicating that the signs are probably too crude to use as instruments to detect varying degrees of latent homosexuality among nonhomosexual males.

Limitations on generalizations from the present results were discussed, and the dangers involved in expecting these signs to serve as unaided indicators of homosexuality in an individual protocol were emphasized. It was concluded that the existing sign schemes must be further refined, validated, and lengthened—a process in which this experiment is only a first step.

REFERENCES

1. ARONSON, M. L. A study of the Freudian theory of paranoia by means of the Rorschach test. *J. proj. Tech.*, 1952, **16**, 397–411.
2. BERGMANN, M. S. Homosexuality on the Rorschach test. *Bull. Menninger Clin.*, 1945, **9**, 78–83.
3. COMMITTEE ON COOPERATION WITH GOVERNMENTAL (FEDERAL) AGENCIES OF THE GROUP FOR THE ADVANCEMENT OF PSYCHIATRY. *Report on homosexuality with particular emphasis on this problem in government agencies.* Topeka: 1955, Report No. 30.
4. DUE, F. O., & WRIGHT, M. E. The use of content analysis in Rorschach interpretation: differential characteristics of male homosexuals. *Rorschach Res. Exch.*, 1945, **9**, 169–177.
5. KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male.* Philadelphia: Saunders, 1948.
6. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique.* New York: World Book Co., 1942.
7. LINDNER, R. M. Content analysis in Rorschach work. *Rorschach Res. Exch.*, 1946, **10**, 121–129.
8. LINDZEY, G., TEJESSY, CHARLOTTE, DAVIDS, A., & HEINEMANN, SHIRLEY. Thematic Apperception Test: a summary of empirical generalizations. Unpublished manuscript, Harvard Univer., 1952.
9. MURRAY, H. A. *Thematic Apperception Test manual.* Cambridge: Harvard Univer. Press, 1943.
10. REITZELL, JEANNE M. A comparative study of hysterics, homosexuals, and alcoholics using content analysis of Rorschach responses. *Rorschach Res. Exch.*, 1949, **13**, 127–141.
11. TEJESSY, CHARLOTTE. The Thematic Apperception Test: a summary and preliminary investigation of empirical generalizations. Unpublished honor's thesis, Harvard Univer., 1952.
12. TOMKINS, S. S. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.
13. WHEELER, W. M. An analysis of Rorschach indices of male homosexuality. *Rorschach Res. Exch.*, 1949, **13**, 97–126.

Received July 7, 1955.

THE EFFECT OF PSYCHOPATHOLOGY ON VISUAL DISCRIMINATION[1]

HERMAN NIEBUHR, JR.
University of Pennsylvania, Medical School
AND DAVID COHEN
VA Hospital, Coatesville, Pennsylvania

THE effect of psychopathology on perceptual processes is relatively unexplored. Inferences about the perceptual functioning of the psychologically ill have generally been drawn from performance on psychological tests. For the most part, such clinical inferences tend to be rather speculative, for behavior on a psychological test is compounded of several functions. Rorschach performance, for example, is not solely a perceptual act, but involves the complex analytical and integrative skills usually associated with thinking. Visual motor performance tests also are compounded of several psychological functions. Consequently, inferences concerning perceptual functioning on these tests are confounded by the presence of extraneous variables and require caution in their evaluation.

The present study of visual-motor processes attempts to determine the contribution of the perceptual component—specifically, visual discrimination—to performance on a well-known clinical instrument, the Bender-Gestalt Test. Bender (1) and others (3, 6), in discussing the functions measured by the Visual Motor Gestalt Test, have placed primary emphasis on the "visual-motor function" per se. Korchin and Basowitz (4) dealt with the perceptual aspects in their tachistoscopic study but again used motor reproduction as their measure of perception. Research on this instrument has in general been directed toward the construction of scoring schemes (2, 5) with little effort to delineate the locus of the distortion one finds in the reproductions of psychopathological subjects. The question may be raised whether patients perceive the designs accurately and then distort them in reproduction, or whether they actually misperceive the design and then more or less accurately reproduce their misperceptions. In the present study, the perceptual aspect of the Bender-Gestalt Test was studied under memory and copy conditions by means of a multiple-choice test. Motor reproduction aspects were also investigated under memory and copy conditions. Comparisons were made between perceptual and motor performance under both conditions.

[1] From the Veterans Administration Hospital, Coatesville, Pennsylvania.

METHOD

Subjects. There were 40 *Ss*: 10 student nurses, 10 acutely schizophrenic patients, 10 chronic schizophrenic patients, and 10 patients hospitalized for neurological reasons. All patients were hospitalized in a Veterans Administration psychiatric hospital at which the student nurses were being trained. Characteristics of the sample are given in Table 1.

Although Pascal and Suttell (5) suggest that age and intelligence are not significantly related to Bender-Gestalt performance, an attempt was made to equate groups for intelligence as far as possible. No attempt was made to control age. All patients were in sufficient contact to cooperate in the experiment, and each *S* completed all tasks.

Materials. Test materials consisted of a standard set of Bender-Gestalt cards and a specially constructed multiple-choice test in which each *S* was required to select from six designs the one which was the exact reproduction of the standard stimulus. A series of alternates was constructed for each of the nine standard designs. In each series, the alternates were of varying similarity to the standard design. No attempt was made to rank the difficulty or similarity of the five wrong choices in each series. Paper cut to the size of the standard Bender-Gestalt stimulus cards was used for the motor reproduction tasks.

Procedure. Each *S* was asked to perform four tasks in the following sequence. There was no delay between these four steps.

I. *Memory Reproduction:* Ten-second presentation of each Bender-Gestalt design from A to 8. After each presentation, *S* was requested to reproduce the design from memory.

II. *Copy Reproduction:* Presentation of each Bender-Gestalt design from A to 8. The *S* was requested to reproduce each design by copying. There was no time limit, and the standard design was kept before *S* until completion.

During the reproduction tasks, each *S* was instructed to reproduce the standard designs as accurately as possible and was cautioned to pay attention to the size of his reproduction, the accuracy of all its parts, and its correct placement on his drawing paper so that his drawings were to be as exact a reproduction of the standard designs as he could possibly effect. The importance of accuracy was reemphasized upon presentation of each succeeding design.

TABLE 1
CHARACTERISTICS OF EXPERIMENTAL AND CONTROL GROUPS

Group	Mean Age	Mean IQ	Mean Time Hospitalization	Diagnosis	N
Student Nurses	19.7	101.5*	—	—	—
Acute Schizophrenic	28.5	102.2**	6.1 mos.	Paranoid	7
				Mixed	2
				Catatonic	1
Chronic Schizophrenic	36.0	96.3**	7.7 yrs.	Paranoid	5
				Hebephrenic	2
				Catatonic	3
Neurological	59.1	Not available	9.9 yrs.	Arteriosclerosis	3
				Multiple sclerosis	1
				Chronic brain synd.	2
				Tertiary syphilis	2
				General paralysis	1
				Epilepsy, grand mal	1

* IQ derived from California Short-Form Test of Mental Maturity, Advanced Form.

** IQ derived from Wechsler-Bellevue Adult Intelligence Scales.

III. *Memory Multiple Choice:* Immediately following the preceding two steps and without further experience with the standard designs, *S* was requested to select each standard design from memory from a series of six alternates for each of the nine designs.

IV. *Matching Multiple Choice:* With each standard design before him, *S* was asked to select the correct design from the same series of six alternates used in Step 3 by matching each alternate with the standard design.

Scoring. For the multiple-choice tasks, a score was obtained for each *S* by tallying the number of correct selections under the memory and matching conditions. The maximum score was 9 and the minimum score was 0 for each condition. For the reproduction tasks, under both memory and copy conditions, each was rated for compliance with the experimental instructions along five general dimensions: size, general placement, angular placement, detail accuracy, and design accuracy. Ratings were made on a 3-point scale, a score of 0 indicating congruence with the original design, a score of 1 indicating minor deviation, and a score of 2 indicating major deviation. Thus, there was a minimum score of 0 and a maximum score of 10 for each design reproduction. For each memory and copy condition, a minimum of 0 and a maximum of 90 points were possible. The higher the score, the less adequate were the reproductions. Specific directions for scoring each of the five variables for each design, as well as scoring examples, were presented to the rater in a Rater Instruction Sheet. A more detailed exposition of the scoring dimensions follows:

I. *Size.* Each design was rated for its size as compared with the original. A reproduction could be larger, smaller, or the same size as the original. A score of 1 was given when the reproduction was slightly larger or smaller than the original. A score of 2 was given when the size was less than half or greater than twice the size of the stimulus.

II. *General Placement.* Each design was rated for its placement on the copy paper in comparison to the placement of the standard design on its card. A score of 1 was given for deviations less than $\frac{3}{4}''$ along either the horizontal or vertical dimensions. A score of 2 was given for deviations greater than $\frac{3}{4}''$ along either or both dimensions.

III. *Angular Placement.* Each design was rated for its angular placement irrespective of its general placement on the copy sheet. A score of 1 was given for a rotation of part or all of the design of less than 45 degrees. A score of 2 was given for rotation over 45 degrees. Thus, severe rotations and reversals were given a score of 2.

IV. *Detail Accuracy.* Each design was rated for the accuracy with which the details of each design were drawn. This did not include the over-all or gestalt accuracy covered by the Design Accuracy score. Since there are different details for each design, the rating was qualitative. Scores of 1 and 2 were given on the basis of the degree of detail distortion. Distortions may take the form of flattening or exaggeration of curves, of drawing loops for dots, dashes for loops or dots, or ellipses for circles, or of increasing or decreasing the size of dots or loops. It is important to recognize that only the components of each design were rated.

V. *Design Accuracy.* Each design was rated for its gestalt quality. This was necessarily a difficult judgment, for it was intended to rate the adequacy of the design as a whole, reflecting both omissions and additions. Slight distortions in the shape of the design, including overlaps and fragmentations, received a score of 1. Major distortions, omissions, and additions received a score of 2. This dimension, as well as the preceding one, was described in special detail in the Rater Instruction Sheet.

Since one of the authors scored each reproduction with full knowledge of each *S*'s group status, and also scored each group by itself, it was necessary to check reliability by having another rater learn the rating system and score the reproductions in random order without knowledge either of *S*'s group membership or of the task instructions. Product-moment reliabilities of .97 and .93 were obtained for designs in the memory and copy conditions respectively. The data presented are based on the scores of the second rater.

RESULTS

Performance on multiple-choice tasks. Tables 2 and 3 present data on the discrimination accuracy between the experimental and control groups on the multiple-choice tasks. An analysis of variance (Table 3) and intergroup comparisons by means of *t* tests indicated the locus and significance of differences among the groups on the perceptual multiple-choice test under memory and matching conditions.

TABLE 2

PERFORMANCE ON MULTIPLE-CHOICE TASKS*

Group	Memory Condition		Matching Condition	
	Mean	Sigma	Mean	Sigma
Student Nurses	5.60	1.78	8.40	.70
Acute Schizophrenic	5.00	1.16	6.80	1.48
Chronic Schizophrenic	3.20	.79	5.50	1.51
Neurological	2.00	1.70	2.90	1.79

* Highest score possible, 9. Lowest score possible, 0.

TABLE 3

ANALYSIS OF VARIANCE OF MULTIPLE-CHOICE TASK SCORES OF ALL GROUPS UNDER MEMORY AND MATCHING CONDITIONS

Source of Variation	*df*	Sum of Squares	*F*	*p*
Between Groups	3	207.25	20.9	.01
Between *Ss*, Same Group	36	119.30		
Total Between *Ss*	39	326.55		
Memory-Matching	1	76.05	104.9	.01
Condition × Group	3	36.85	16.9	.01
Pooled *Ss* × Condition	36	26.10		
Total Within *Ss*	40	139.00		
Total	79	465.55		

Despite the small number of *Ss* in each group, statistical significance between groups under the memory condition is of a high order with but one exception—the difference between the student nurses and the acute schizophrenics ($p = <.40$). Differences between groups under the matching condition are again all significant with but one exception—the difference between the acute and chronic schizophrenics. The student nurses showed the greatest improvement from the memory to the matching condition, with 5 out of the 10 *Ss* obtaining perfect scores on the matching condition. No *Ss* in the three patient groups equalled this performance despite the fact that cooperation seemed high in all groups.

Performance on reproduction tasks. Consistent differences among the groups in the adequacy of reproduction are apparent from Table 4. Table 5 summarizes the results of an analysis of variance. Again, the differences between groups as studied by *t* tests are for the most part statistically significant. The difference between the nurses and the acute schizophrenics under the memory condition approaches, but does not reach, significance ($p < .20$). This is also the case in the comparison of the acute and chronic patients ($p < .10$). Under the copy condition, all groups differed significantly except for the difference between the acute and chronic schizophrenic groups ($p < .20$). All nonsignificant differences were in the expected direction. Again, the nurses improved the most from the memory to the copy condition.

A series of *t* tests was made among the groups of the five component scores composing the total reproduction score. In general, it was found that differences between the groups on the reproduction score under the copy condition were greater than under the memory condition. However, while the nurses differed significantly from the chronic schizophrenic and neurological groups on almost all of the five component dimensions, there were no significant differences on any dimension between the control and the acute schizophrenic groups on the memory condition. Under the copy condition, the nurses differed significantly from the acute schizophrenics as well as from the other groups on all dimensions except the angular placement dimension. Similarly, the

TABLE 4

PERFORMANCE ON REPRODUCTION TASKS*

Group	Memory Condition		Copy Condition	
	Mean	*SD*	Mean	*SD*
Student Nurses	21.6	8.87	9.80	4.19
Acute Schizophrenic	28.0	10.60	21.8	6.86
Chronic Schizophrenic	37.4	9.16	28.0	12.15
Neurological	60.4	10.26	50.8	11.35

* Best score, 0. Worst score, 90.

TABLE 5

ANALYSIS OF VARIANCE OF REPRODUCTION TASK SCORES OF ALL GROUPS UNDER MEMORY AND COPY CONDITIONS

Source of Variation	*df*	Sum of Squares	*F*	*P*
Between Groups	3	17,216.95	30.5	.01
Between *Ss*, Same Group	36	6,782.00		
Total Between *Ss*	39	23,998.95		
Memory-Copy	1	1,461.25	112.2	.01
Condition × Group	3	329.05	8.4	.01
Pooled *Ss* × Condition	36	469.70		
Total Within *Ss*	40	2,260.00		
Total	79	26,258.95		

acute schizophrenic group differed significantly from the neurological group on all dimensions, while the chronic schizophrenic group differed significantly from the neurological group on three of the five dimensions. There were no significant differences between the two schizophrenic groups on any of the dimensions. Thus, four of the five component dimensions provided a fair degree of discrimination between the groups. A combination of all five dimensions, however, achieved the greatest discrimination and is thus likely to prove most useful in clinical application.

Comparison of multiple-choice and reproduction tests. It seemed useful to analyze the relationship between perceptual accuracy as measured by the multiple-choice test and motor performance as measured by the reproduction score. Within-group correlations were with one exception positive but nonsignificant.

DISCUSSION

The results clearly indicate that perceptual efficiency, insofar as visual discrimination is concerned, decreases with severity of psychopathology. The adequacy of motor reproduction likewise decreases with severity of psychopathology. The question of the degree and nature of the relationship between perceptual and motor efficiency remains unresolved on the basis of the present data. The positive, although nonsignificant, correlations between the two functions within each group leave the matter ambiguous at this time. However, it is to be noted that age and severity of psychopathology are highly correlated in the experimental population. Hence, a decrement in both perceptual and motor functioning as a function of age is a possibility which cannot be ruled out in spite of Pascal and Suttell's (5) results.

While the memory condition was included to impose an additional burden for the patient groups and thus to insure greater differences from the nurses, it was evident that the additional pressure of the memory condition served only to decrease those differences. The inferior performance of the patient groups under the matching condition, in which there was no time limit and the comparison design was constantly present, only serves to emphasize the marked differences in perceptual functioning. The superiority of the acute schizophrenics over the chronic schizophrenic group is consistent with Zubin's (7) finding that patients whose perceptual efficiency is less impaired have a better prognosis for hospital discharge. Thus, the evaluation of a patient's perceptual functioning as a prognostic aid appears to be a fruitful area for further investigation.What the reason is for the decrement in performance in patient groups can only be speculated upon. Whether it represents a regressive phenomenon in the sense of a return to an earlier mode of perceptual development, an impairment of the attention and concentration requisite to the accurate perception of a design, or some other process, cannot be determined from this study.

The differential levels of performance of the experimental and control groups on the reproduction tasks concur with earlier clinical and experimental studies with the Bender-Gestalt Test. The methodological modifications made in the present study seem to indicate additional useful dimensions. Not only were the detail and design accuracy dimensions, like those used in other Bender scoring systems, discriminating between the groups, but the dimensions of placement and size were also useful discriminators. The use of paper of the same size as the stimulus cards seemed to provide a more accurate frame of reference against which to evaluate performance.

A refinement of this approach is planned so as to explore further the specific nature of the perceptual disturbance involved in psychopathology.

SUMMARY

Perceptual and motor aspects of the Bender-Gestalt Test performance were compared for four groups of *Ss*: 10 student nurses, 10 acute schizophrenic patients, 10 chronic schizophrenic patients, and 10 patients hospitalized for neurological reasons. All patients were hospitalized in a psychiatric hospital at which the student nurses were being trained. After the *Ss* had obtained equal experiences with the Bender designs by reproducing them under memory and copy conditions, a perceptual multiple-choice test was administered under memory and matching conditions. Results indicate that there are marked differences in perceptual accuracy between the four groups: the control population was most accurate, with

the acute schizophrenic, chronic schizophrenic, and neurological patients less proficient in that order. Statistical significance was obtained on all but two comparisons. The results of the motor reproduction performance are similar to perceptual results.

The question of the degree and nature of the relationship between perceptual and motor efficiency remains unresolved in this study. Since age and severity of psychopathology are highly correlated in the experimental populations, one cannot rule out the possibility that decreasing perceptual and motor efficiency is a function of age in spite of Pascal and Suttell's (5) results.

REFERENCES

1. BENDER, LAURETTA. A visual motor test and its clinical use. *Res. Monogr. Amer. Orthopsychiat. Ass.*, 1938, No. 3.
2. BILLINGSLEA, F. Y. The Bender Gestalt: an objective scoring method and validating data. *J. clin. Psychol.*, 1945, **1**, 134–140.
3. HALPERN, FLORENCE. The Bender Visual Motor Gestalt Test. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 324–340.
4. KORCHIN, S., & BASOWITZ, H. The tachistoscopic Bender-Gestalt test. Paper read at Amer. Psychol. Ass., New York, Sept., 1954.
5. PASCAL, G. R., & SUTTELL, B. J. *The Bender-Gestalt Test: quantification and validity for adults*. New York: Grune & Stratton, 1951.
6. WOLTMANN, A. The Bender Visual-Motor Gestalt Test. In L. E. Abt & L. Bellak (Eds.), *Projective psychology*. New York: Knopf, 1950. Pp. 322–356.
7. ZUBIN, J., WINDLE, C., & HAMWI, V. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, **21**, 342–355.

Received August 12, 1955.

THE ROTATION OF DRAWINGS BY BRAIN-DAMAGED PATIENTS[1]

AUBREY J. YATES

Institute of Psychiatry, Maudsley Hospital, London

THE block design rotation effect was first observed in psychiatric patients on the Goldstein Cube test, which requires the *S* to reproduce patterns with one-inch colored cubes. Some patients, while completing the designs correctly, left the completed pattern in a rotated position, apparently without awareness of the rotation (6). Such rotations frequently reached 45° (Fig. 1) but rarely exceeded it.

In previous work carried out by Shapiro (6, 7, 8), an attempt was made to determine the factors which influence the appearance of the rotation effect in both normal and abnormal *S*s and to discover the factors which facilitate its appearance in exaggerated form in certain *S*s.

The principal findings to date are of two kinds. First, brain-damaged psychiatric patients rotate significantly more than non-brain-damaged psychiatric patients (7); and second, the rotation effect appears under specifiable conditions. For example, when the line of symmetry of a design (defined as the line which divides the design into mirrored halves) is at an angle to the vertical axis of the total visual field, the tendency to rotation is increased. When the line of symmetry is parallel to the vertical axis of the total visual field, then the tendency to rotation is decreased. Also, when the design is in a diamond orientation, the tendency to rotation is increased; but when the design is in a square orientation, the tendency to rotation is decreased. This orientation is referred to as the figure shape. Finally, when the ground (defined as the 6″ by 6″ white card on which the design is placed) is in a diamond orientation, the tendency to rotation is increased; when the ground is in a square orientation, the tendency to rotation is decreased.

Thus, brain-damaged psychiatric patients tend to produce the rotation effect and to do so under identifiable conditions. Shapiro (8) explained these results in terms of an exaggeration of negative induction effects in brain-damaged patients. The concept of negative induction used is that developed by Pavlov (5), i.e., the notion that the development of excitatory effects in any part of the nervous system is followed or accompanied by inhibitory effects in other parts. The inhibitory aspects of negative induction are assumed to be exaggerated in brain-damaged patients. The justification for this generalization is found mainly in the work of Bender (1) and his associates.

From this reasoning, two predictions were made: First, brain-damaged patients would rotate significantly more than non-brain-damaged patients in an analogous situation. They were therefore required to draw the designs instead of making them with blocks. Such a procedure is much shorter and is considerably easier for the *S*; it is also less expensive to record than previous methods, which involved the use of a camera. Second, the factors influencing rotation in the blocks test would be the same in the drawing version, i.e., a diagonal line of symmetry would induce more rotation than a vertical, a diamond figure more than a square, and a diamond ground more than a square. The experiment reported was designed to test these predictions.

METHOD

Normal control subjects (N = 25). The nonpsychiatric control *S*s consisted of a group of female student nurses. Their mean age was 20.00 (*SD*, 0.58; range, 19–23), and mean equivalent Wechsler[2] full-scale IQ, 116.24 (*SD*, 10.58; range, 92–142).

Non-brain-damaged psychiatric patients (N = 25). The psychiatric group included patients at various stages of illness and of various degrees of acuteness. In terms of psychiatric diagnosis, it included six cases of

[1] The writer expresses his appreciation to Dr. M. B. Shapiro for his helpful criticisms in the design of the experiment. Originally part of a thesis submitted to the University of London in partial fulfillment of the requirements for the degree of doctor of philosophy, this work was made possible by a grant from the Research Fund, made available from the endowment by the Board of Governors of the Bethlem Royal Hospital and the Maudsley Hospital.

[2] The equivalent Wechsler full-scale IQ was calculated by summing the WS for Vocabulary, Similarities, and Block Design and then prorating. This combination of subtests has been shown to correlate +0.937 with full-scale IQ in normal *S*s (4) and +0.900 in a random sample of 100 psychiatric patients (9).

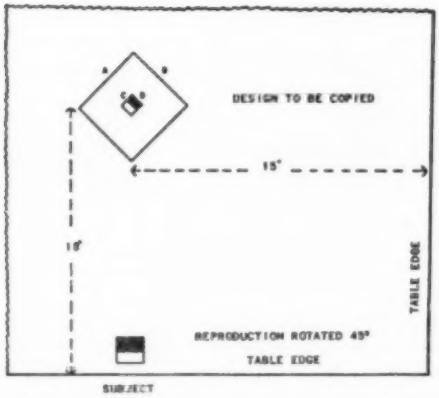

FIG. 1. EXAMPLE OF A CARD FROM THE BLOCK DESIGN ROTATION TEST WITH THE REPRODUCTION ROTATED 45°

schizophrenia, seven of depression, two of obsessional neurosis, six of anxiety neurosis, three of conversion hysteria, and one of mixed neurosis. The mean age was 28.96 (*SD*, 6.25; range, 17–45) and the mean equivalent IQ 118.08 (*SD*, 16.89; range, 76–144). There were 15 males and 10 females.

Brain-damaged patients (N = *20*). The brain-damaged group included 12 patients with tumors, three with posttraumatic syndrome, one with aqueduct stenosis, and one with cerebral atrophy. The remaining three cases were tested after temporal lobectomy for psychomotor epilepsy. The mean age of this group was 42.35 (*SD*, 7.98; range, 26–52), and the mean equivalent IQ 104.15 (*SD*, 16.61; range, 76–139). There were 14 males and six females.

Tests. The tests employed were three subtests of the Wechsler Intelligence Scale, Form I (Vocabulary, Similarities, and Block Design), and the Drawing Version of the Block Design Rotation Test. For the new version of the test, Shapiro's (7) 40 designs were used. Instead of copying each design with four Kohs blocks, however, the *S* was required to draw it in an exercise book provided for this purpose and placed in a fixed position relative to the design. Several methods of measuring rotation were tried and rejected for various reasons. Eventually, it was decided to measure the angle of the *first* side drawn by the *S*.[3]

RESULTS

Association of the rotation phenomenon with brain damage. Analysis of variance and covariance carried out on the three groups for rotation and intelligence (Table 1) showed that the brain-damaged group rotated significantly more than the other two groups (at the .03 level of confidence, using a one-tailed test) when the effect of differences in level of intelligence[4] between the groups was held constant. These results may be compared with those obtained for three very similar groups using the block design version of the test (Table 1). The rotation effect with Kohs blocks discriminated at a much higher level of confidence (when intelligence was held constant) than was the case with the drawing version. It is clear, therefore, that while the results support the hypothesis that the rotation effect is associated with brain damage, the drawing version is a relatively insensitive measure of it compared with the blocks. In its present form, the overlap between groups in the drawing version is too large for the test to have much clinical usefulness.

There was no relationship between rotation and age or sex in any of the groups.

Factors influencing the appearance of the rotation effect. The principal results of an

TABLE 1

ANALYSIS OF VARIANCE AND COVARIANCE
(Block design version contrasted with drawing version)

		Rotation (Y)		Intelligence (X)	
	N	M	σ	M	σ
Blocks Version					
Brain damaged	38	8.02	4.83	104.11	16.80
Functional	19	2.86	1.60	108.37	12.90
Control	20	4.42	3.05	113.85	13.50
Analysis of variance		$F = 13.05$ $p = <.01$		$F = 2.70$ $p =$ NS	
Analysis of covariance		$F_{Y \cdot X} = 11.86$ $p = <.01$			
Drawing Version					
Brain damaged	20	6.85	5.16	104.15	16.61
Functional	25	4.10	2.64	118.08	16.89
Control	25	3.31	1.04	116.24	10.58
Analysis of variance		$F = 7.17$ $p = <.01$		$F = 5.64$ $p = .03$	
Analysis of covariance		$F_{Y \cdot X} = 3.38$ $p = .03$			

There was no relationship between rotation and age or sex in any of the groups.

Factors influencing the appearance of the rotation effect. The principal results of an

[3] Full details of the method of administering and scoring the drawing version of the test are contained in Yates (9).

[4] The relationship between rotation scores and intelligence level is complex and will be the subject of a separate paper, in which a large number of groups is examined.

TABLE 2
ANALYSIS OF VARIANCE OF ROTATION—*F* RATIOS AND SIGNIFICANCE LEVELS

Variable	Group					
	Control		Functional		Brain Damaged	
	F	*p*	*F*	*p*	*F*	*p*
Figure (F)	434.201	<.01	114.191	<.01	131.145	<.01
Ground (G)	1.954	NS	7.098	<.01	57.195	<.01
Persons (P)	5.030	<.01	15.312	<.01	17.946	<.01
F × P	9.482	<.01	11.629	<.01	14.280	<.01
G × P	—	NS	2.447	.01	10.669	<.01
F × G	—	NS	8.851	<.01	1.400	NS
P × G × F	—	NS	2.691	.01	3.560	.01

analysis of variance of the factors thought to influence rotation for each of the three groups are shown in Table 2. *F* ratios are not shown where the mean square for the variable was less than the residual mean square.

Contrary to prediction, there was no evidence that the line of symmetry affected rotation. Previous analyses of the data from experiments using the blocks version of the test revealed consistently from group to group that the angle of the line of symmetry in combination with the figure shape was the most important single variable influencing the appearance of rotation. When these two factors were in agreement in terms of directional value in space, rotation was maximized if both were at an angle but minimized if both were vertical. In the drawing version, however, this significant effect of the line of symmetry on rotation was not found. While the angled line of symmetry appeared to produce slightly more rotation than the vertical line of symmetry, the difference was always small and insignificant.

The most significant factor influencing rotation in the drawing version was figure shape alone. When the figure shape is square, rotation is minimal; when the figure shape is diamond, rotation is maximal, *irrespective of the direction of the line of symmetry*. In the control and psychiatric groups, the line of symmetry played no role whatever in producing the rotation effect; in the brain-damaged group, although it just attained significance level, its effect was very small indeed when compared with the influence exerted by the figure shape.

The *S*s differed significantly with respect to the *relative* influence of square and diamond

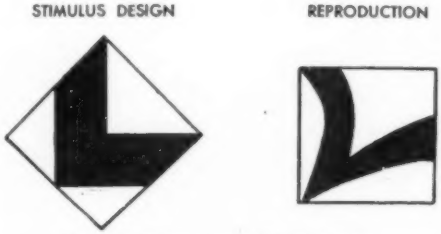

FIG. 2. EXAMPLE OF A DISTORTED OR "TWISTED" REPRODUCTION

figure shape. A much higher proportion of the total rotation score was produced by the diamond figure in some *S*s than in others. This is shown by the F × P interaction in Table 2.

A diamond-shaped ground induced significantly more rotation than a square-shaped ground in the functional and brain-damaged groups, but not in the control group. The effect of ground shape was, however, relatively weak as compared with the influence of the figure shape.

Another result was quite unexpected. A number of the brain-damaged *S*s first of all drew the figure shape in a rotated position. They then drew the inner content in a correct orientation (with respect to the stimulus object) but apparently without reference to the figure shape which they had just rotated. Sometimes, a form of compromise appeared to result and a distorted or "twisted" reproduction was drawn (Fig. 2). In some instances, the figure shape was distorted, but not the line of symmetry; in other instances, the reverse was the case.

DISCUSSION

It seems clear that copying designs by drawing produces essentially similar results as reproducing them by manipulating blocks. This result is consistent with other findings. Hanvik and Andersen (3), for example, reported that 59 per cent of a brain-damaged group of 44 patients produced one or more rotations of at least 30° while drawing the figures of the Bender-Gestalt test. Similarly, Hanvik (2) reported an 80 per cent incidence of abnormal EEG records in a group of 20 children showing rotations of one or more of the Bender figures.

Another finding of importance was the lack

of effect in the drawing version of the angle of the line of symmetry. Coupled with this is the observation that certain brain-damaged patients produced distorted or "twisted" copies of the designs. These two results may possibly be linked together. Nearly all *S*s copied the designs by first of all drawing the outline, i.e., the figure shape. They then proceeded to complete the content or "inside" of the design. Thus, it was possible for an *S* to complete the outline without reference to the content and vice versa. In some brain-damaged patients, this is exactly what seemed to happen. In some instances, a kind of compromise resulted and a distorted or "twisted" design was drawn. In the blocks version, on the other hand, the situation is typically quite different. Here, the "content" and figure shape change at the same time, whichever aspect of the design is responded to. Thus, if the *S* rotates the figure shape, then the line of symmetry also rotates.

In general, the production of such "twisted" drawings may be accounted for in the same terms as the rotation effect. The brain-damaged patient may suffer from exaggerated negative induction effects in the sense of ignoring surrounding perceptual cues. In this instance, having rotated the figure shape, he then draws the content without reference to these cues. It should be pointed out that this explanation has been derived from a consideration of the results of the various experiments carried out on the rotation effect. It is therefore in the nature of a plausible hypothesis which remains to be tested.

SUMMARY

1. Rotation effects in a drawing version of a task requiring the copying of designs discriminated between brain-damaged patients on the one hand and non-brain-damaged psychiatric patients and normal control *S*s on the other (but to a less satisfactory degree than the blocks version of the test).

2. The relative influence of the factors affecting the appearance of the rotation effect in all *S*s is shown to be different in the drawing version than in the blocks version. There was no influence from the line of symmetry. Figure shape exerted the main influence.

3. A tentative explanation of the disappearance of the influence of the line of symmetry and of a new finding of distorted or "twisted" reproductions is put forward.

REFERENCES

1. BENDER, M. B. *Disorders in perception.* Springfield, Ill.: Charles C Thomas, 1951.
2. HANVIK, L. J. A note on rotations in the Bender-Gestalt test as predictors of *EEG* abnormalities in children. *J. clin. Psychol.*, 1953, **9**, 399.
3. HANVIK, L. J., & ANDERSEN, A. L. The effect of focal brain lesions on recall and on the production of rotations in the Bender-Gestalt test. *J. consult. Psychol.*, 1950, **14**, 197–198.
4. HILDEN, A. H., TAYLOR, J. W., & DUBOIS, P. H. Empirical evaluation of short Wechsler-Bellevue scales. *J. clin. Psychol.*, 1952, **8**, 321–331.
5. PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1946.
6. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. I. Initial experiments. *J. ment. Sci.*, 1951, **97**, 90–110.
7. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. II. Confirmatory and explanatory experiments. *J. ment. Sci.*, 1952, **98**, 605–617.
8. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. III. The testing of an explanatory theory. *J. ment. Sci.*, 1953, **99**, 394–409.
9. YATES, A. J. An experimental study of the block design rotation effect with special reference to brain-damage. Unpublished doctor's dissertation, Univer. of London, 1954.

Received October 3, 1955.

HOPE OF SUCCESS AND FEAR OF FAILURE AS ASPECTS OF NEED FOR ACHIEVEMENT[1]

RUSSELL A. CLARK, RICHARD TEEVAN,
University of Michigan
AND HENRY N. RICCIUTI
University of Colorado School of Medicine

RESEARCH by McClelland and his associates (3) on a group TAT measure of need for achievement (n Achievement) yielded evidence suggesting that this measure involves two recognizable aspects of achievement motivation, *hope of success* (HS) and *fear of failure* (FF). One is an approach motive involving anticipation of reward, the other an avoidance motive involving anticipation of punishment. The main source of evidence for this distinction (summarized in [3]) consists of the repeated finding that individuals with moderate or low n Achievement scores appear fearful or defensively oriented whereas individuals with high scores appear hopeful. For example, those with moderate or low scores inhibit the recall of tasks failed, and are slow to recognize the tachistoscopic presentation of achievement-related words connoting failure, obstacles to achievement, etc. Subjects with high n Achievement scores, on the other hand, are especially quick in the recognition of achievement-related success words and show none of the defensive characteristics of the moderate group.

Clarification of these suggested relationships requires the measurement of fear of failure and hope of success, and several attempts have been made to arrive at such measures. The senior investigator first sought to arouse fear of failure and hope of success experimentally, administering the group TAT to subjects after they had been led to anticipate doing either very well or very poorly on the examination that they were about to take. Results failed to support the expectation that certain of the subcategories composing the total n Achievement score, singly or in combinations, would distinguish between these two experimentally different groups.

Another attempt involved a factor analysis of the intercorrelations among these subcategories in the hope that fear of failure and hope of success might emerge clearly in such an analysis, and that categories or combinations of categories proving to be heavily saturated with the two respective factors could then be used for indices. Several such analyses were made involving somewhat different categories, combinations of categories, and criterion measures. This approach in turn was unsuccessful.

A fresh approach to the problem, a measure of these two aspects of n Achievement, was therefore sought that was independent of the measurement of the achievement motive. Studies of level of aspiration seemed to offer a way of isolating the cautious, defensive individual (FF) as opposed to the hopeful person (HS). A questionnaire on level of aspiration was therefore designed for the present purpose.

Because of recognized complexities in the measurement of level of aspiration, it had been given low priority in our previous approach to the problem. Our experience with such a measure seems worth reporting, however, not only because it yielded some positive results but also because it shed light on one plausible reason why some of the previous attempts had failed.

METHOD

The group TAT measure of n Achievement had been administered to three classes of Swarthmore freshmen under so-called neutral conditions about five weeks prior to the final examination. The level-of-aspiration questionnaire was administered to these classes immediately before the examination. It read as follows:

> You are asked to cooperate in a study of students' attitudes toward examinations by answering several questions which are explained below. The information you record will be held in strict confidence, and *will not* be seen by your instructors, or by anyone concerned with your course grades. Please express yourself as honestly and candidly as you can since the success of the study will depend upon your giving your frank and best opinion.
>
> *Question 1.* On the grade scale in the table below,

[1] This study was supported by a grant from the Office of Naval Research under the supervision of David C. McClelland.

you are to indicate the two grades between which you are *practically certain* your final examination grade will fall. Make this as objective a judgment as you can, following these directions.

Lower Extreme: The lower grade you mark would answer this question: "I'm practically certain I will get at least a ——."

Upper Extreme: The higher grade you mark would answer this question: "I'm practically certain I won't get a better grade than a ——."

Grade Scale

E D− D D+ C− C C+ B− B
B+ A− A A+

Question 2. For *each* of the grades in the table above, indicate how you believe you would feel if you made that particular grade on your final examination. Do this by placing *below each grade from E to A+*, the number of the most appropriate statement below, which would describe how you would feel. Do this on the line marked "Question 2."

+3 I would feel extremely good and very satisfied with my grade.
+2 I would feel quite good and quite satisfied.
+1 I would feel somewhat satisfied.
0 I would feel about neutral—neither good nor bad.
−1 I would feel somewhat disappointed.
−2 I would feel quite disappointed and dissatisfied.
−3 I would feel extremely bad, and very much dissatisfied with my grade.

Question 3. Supposing that it were possible for you to be excused from taking the final examination you are about to take. In that case, would you be willing to settle for an A+? For *each of the remaining grades* from E to A+, indicate whether you *would* (Y) or *would not* (N) be willing to settle for that grade rather than take the examination, if that were possible. Do this by marking a Y or N under each grade.

Question 4. What grade are you really going to try to make on the examination?

Thank you very much for your cooperation. When you have finished, please put this sheet in the envelope and seal the envelope.

In the questionnaire, Question 1 calls for a realistic estimate of the limits within which the students' grades will fall. Question 2 calls for the *affect* that would be experienced upon receiving any particular grade. Question 3, which is somewhat of an innovation in level-of-aspiration studies, asks for the lowest grade that the student would "settle for" in lieu of taking the final exam. The fourth question is one frequently asked in level-of-aspiration studies. It asks for the grade that the individual is "really going to try to make."

How can these items be used to distinguish FF from HS? Several possibilities were entertained. First, we expected that an FF student might settle for a grade that was farther below his upper realistic limit than would an HS person (Index I). Such a defensive, fearful person should accept a lower grade in relation to his realistic expectations because doing so minimizes possibility of failure. An HS student, however, is optimistic and should be unwilling to settle for a grade very far removed from his realistic upper limit. An equally meaningful index would take into consideration the range of expected scores. A second index therefore, was obtained by dividing Index I by the range. Still another index follows from the consideration that, other things being equal, the affect associated with the grade "settled for" should be more positive for the HS person than for the FF person. The HS person should not settle for a grade unless he is fairly well satisfied with it, while an FF individual should be willing to compromise in this respect. A fourth index is related to the third, and attempts to correct for differences in realistic expectations. The difference between the affect associated with the upper limit grade and the grade "settled for" should be less for the HS group than for the FF group. For example, an HS person who realistically expects to get a poor grade might settle for a grade close to this level without being highly pleased with it, provided that he anticipates that it would give almost as much satisfaction as the upper realistic limit. This formulation permits HS subjects to have low or negative absolute levels of affect, but requires that these levels be close to the level of affect associated with the upper limit.

In treating responses to Question 1, adjacent grades were considered to be one unit apart. The distribution of scores for Index I (Upper Limit minus Settle For) was grouped into seven class intervals, with a value of seven being assigned to the smallest difference. Since the prediction is that for the HS person the quantity (Upper Limit minus Settle For) will be small, the data were coded so that HS subjects would receive the highest scores. The results for the other three indices were treated in parallel fashion.

RESULTS

Table 1 presents intercorrelations showing the internal consistency of these indices. It shows that the correlations are all positive and highly significant. For establishing further relations with the HS–FF continuum all four categories were summed in order to obtain a more stable though crude measure. This distribution was divided into thirds with the

TABLE 1

RELATIONSHIP AMONG FOUR INDICES OF HOPE OF SUCCESS AND FEAR OF FAILURE*

(Each index is correlated with the sum of the other three)

Index	Correlation with Remaining Three Indices
Upper Limit minus Settle For	.85
Upper Limit minus Settle For Range	.82
Affect over Settle For	.58
Affect over Upper Limit minus Affect Settle For	.83

* For $N = 45$ a correlation of .38 is significant at the .01 level of confidence.

TABLE 2

RELATIONSHIP BETWEEN THE HOPE OF SUCCESS-FEAR OF FAILURE CONTINUUM AND QUESTIONNAIRE RESPONSES

(χ^2 test of independence)

Questionnaire Items	χ^2*	p
Settle For	27.45	$<.01$
Level of Aspiration	9.12	.06
Upper Limit	1.57	—
Lower Limit	8.20	$<.10$

* χ^2 was calculated using the method of maximum likelihood.

TABLE 3

TOTAL N ACHIEVEMENT SCORES FOR THE HOPE OF SUCCESS, MIDDLE, AND FEAR OF FAILURE GROUPS

(F_1 = between groups/within groups; F_2 = Middle and combined HS, FF group/within groups)

Group	N	Mean n Achievement Score
HS	14	11.57
Middle	16	18.00
FF	15	10.73
	$F_1 = 4.26$	$F_2 = 8.44$
	$p < .05$	$p < .01$

high third showing hope of success in clearest form and the low third showing the most consistent fear of failure. The middle third might be characterized either in terms of conflict or of realistic orientation, alternatives that we are not at present in a position to distinguish.

Table 2 presents the relationship between this three-way categorization and four separate questionnaire responses: Upper Limit, Lower Limit, Settle For, and Level of Aspiration. We had made no prior predictions concerning these relationships, but, other things being equal, one would expect that the HS subjects would have higher absolute levels of Settle For and Level of Aspiration. In our prediction we had used Settle For, but the contention was that it would be closer to the Upper Limit for the HS students and not necessarily at a higher absolute value. In any event both the Settle For and Level of Aspiration turn out to be higher for the HS than for the FF subjects.

We had no expectations about the relationship with Upper Limit and Lower Limit. The HS group and the Middle group have about the same distribution at upper and lower limits, whereas the FF group is lower in Upper and Lower Limit. The HS group and the Middle group have about the same distribution at upper and lower limits, whereas the FF group is lower in upper and lower limits than either of the other groups.

Turning now to the relationship with the TAT measure of n Achievement let us first examine the relationship between the HS–FF continuum and the total n Achievement score. Table 3 presents an analysis of variance of this threefold relationship. It is quite clear that the Middle group has the highest scores with HS and FF groups scoring at about the same level. In previous work persons with high scores had showed HS characteristics, with moderate or low scores being associated with signs of fear of failure. The present results are in obvious disagreement.

With respect to subcategories in the measurement of n Achievement, the McClelland scoring system includes several clearly positive and negative categories, which are given below. The positive categories represent what McClelland has earlier termed "Goal Imagery"; the negative categories represent what he has called "Deprivation Imagery."

Goal Imagery

Positive Categories

I+ Successful Instrumental Activity
GA+ Anticipations of Success
G+ Positive Affect over Goal Achievement

Deprivation Imagery

Negative Categories

I− Unsuccessful Instrumental Activity
GA− Anticipation of Failure
G− Negative Affect over Failure to Achieve Goal
B Blocks or Obstacles in Path to Achievement

The following analyses involve the use of positive and negative subtotals computed from the scored protocols of each individual.

Table 4 presents an analysis of variance for both the positive and negative subscores. The Middle group employs significantly more negative categories than either the HS or FF groups. In regard to the positive subscores the three group means are very similar. The group differences in total n Achievement score are thus accounted for mainly in terms of the negative categories.

Because the Middle group has a much higher number of achievement-related stories, however, this group should tend to show a

TABLE 4

POSITIVE AND NEGATIVE n ACHIEVEMENT SUBSCORES FOR THE HS, MIDDLE, AND FF GROUPS
(F_1 = between groups/within groups; F_2 = middle and combined HS, FF group/within groups)

Groups	Mean n Achievement Subscores		
	N	Negative	Positive
HS	14	1.78	1.43
Middle	16	3.81	1.37
FF	15	1.87	1.47
	$F_1 = 4.22$ $p < .025$	$F_2 = 8.43$ $p < .01$	F_1, n.s.; F_2, n.s.

TABLE 5

MEDIAN ANALYSIS OF HS–FF IN RELATION TO FREQUENCY OF POSITIVE AND NEGATIVE CATEGORIES

Group	Number of Subjects above or below Median with respect to:			
	Negative Categories		Positive Categories	
	Above median	Below median	Above median	Below median
HS	6	8	10	4
Middle	11	5	4	12
FF	11	9	9	6
	$\chi^{2*} = 3.13$ $p = .20$		$\chi^{2*} = 7.41$ $p = .03$	

* χ^2 was calculated using the method of maximum likelihood.

higher frequency for all categories. To take account of this difference the total positive and negative scores for each individual were divided by the number of his achievement-related stories. Table 5 presents the results of a median test for the number of positive and negative categories per achievement-related story. The median test is appropriate to the present data since it makes no assumptions about the variances, which would be dubious in the case of ratio scores. These results show that the HS and FF groups are about equal in relative frequency of positive categories, but are significantly higher than the Middle group. The findings for relative negative frequency are just the reverse except that the differences are not highly reliable.

DISCUSSION

Let us first examine the treatment of this general problem offered by McClelland (2).

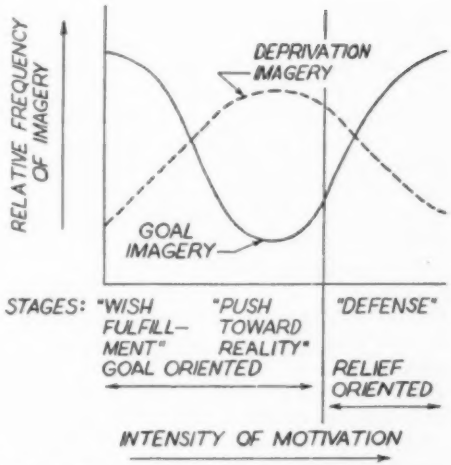

FIG. 1. THE HYPOTHETICAL EFFECT OF INCREASED MOTIVE INTENSITY ON THOUGHT PROCESSES

Reproduced by permission from McClelland (2). Copyright by William Sloane Assoc.

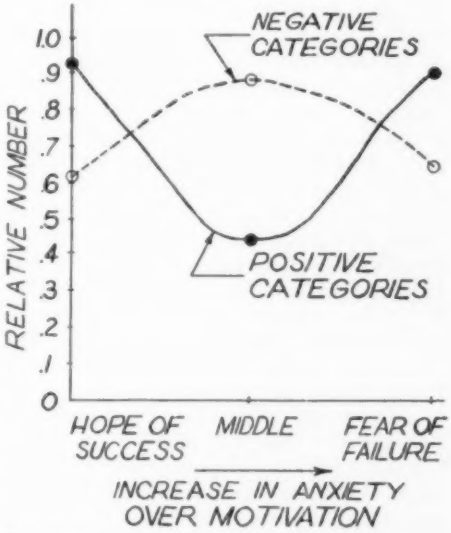

FIG. 2. THE NUMBER OF POSITIVE AND NEGATIVE CATEGORIES PER ACHIEVEMENT-RELATED STORY AS A FUNCTION OF THE HS–FF CONTINUUM.

Because there were five negative categories and only three positive categories used in this analysis, the number of positive categories has been multiplied by five-thirds in order to make the two curves directly comparable.

After a thorough review of the evidence related to the effects of different levels of intensity of motivation, McClelland presents the hypothetical functional relationships which are reproduced in Fig. 1. It can be seen that these curves are very similar to those in Fig. 2 which represent a plot of the results obtained from the analysis of the number of positive and negative categories per achievement-related story. McClelland's "Wish Fulfillment" stage, which corresponds to our HS group, represents a condition of satisfaction and no anxiety over failure to attain the goal in question. His "Push Toward Reality" stage corresponds to what we have called the Middle group. Here there is decreasing optimism and satisfaction and a moderate increase in anxiety over goal attainment. His final "Defense" stage is analogous to our FF group. Here anxiety is so great and hope of goal attainment so small that the individual must repress deprivation-imagery stimuli and enhance goal-related imagery. The interested reader should consult McClelland (2) for the complete treatment of this problem.

One general implication strongly suggested by this study is that the level of n Achievement score *cannot* be used as an index of the FF–HS continuum. It is quite probable that the total n Achievement scores obtained by groups along this continuum are a function of the pictures used to obtain the n Achievement scores. The discrepancy between the present results and earlier results could easily result from the fact that of the eight pictures used in this study, only one was the same as that used in the earlier work.

The present research also indicates why some of the earlier work failed to distinguish fear of failure and hope of success as aspects of achievement motivation. The present research shows that FF subjects are *phenotypically* similar to subjects characterized by Hope of Success. With regard to the TAT measure of n Achievement the total score and the positive and negative subscores are similar for both of these extreme groups. It is the inclusion of the Middle group that gives rise to significant differences. In the initial attempt to arouse hope of success and fear of failure experimentally, successful arousal would lead only to the two extreme groups. Since no Middle group was provided for, the expected differences could not be found.

SUMMARY

Previous research has suggested that McClelland's TAT measure of n Achievement involves two recognizable aspects—hope of success (HS) and fear of failure (FF). A level-of-aspiration questionnaire related to grades on a course examination was designed to give an independent measure of the HS–FF continuum, and administered to college students in conjunction with the examination. Relations between this measure and both total n Achievement score and various n Achievement subscores were investigated. In brief, the findings were these: (*a*) students at the extremes of the continuum have lower n Achievement scores than students in the middle of the continuum. (*b*) in terms of positive subscores (goal imagery) on the n Achievement measure, the extremes have higher scores than subjects in the middle of the HS–FF continuum. (*c*) In terms of negative subscores (deprivation imagery) the Middle group tends to score more highly than the extremes.

REFERENCES

1. CLARK, R. A. The projective measurement of experimentally induced levels of sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391–399.
2. MCCLELLAND, D. C. *Personality*. New York: Sloane, 1951.
3. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts. 1953.
4. MOOD, A. M. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1950.

Received September 22, 1955.

SEXUAL SYMBOLISM AND THE VARIABLES OF SEX AND PERSONALITY INTEGRATION[1]

AUSTIN JONES[2]

Central State Hospital, Indianapolis

THE psychoanalytic theory of sexual symbolism has been notably lacking in experimental support since its formulation by Freud in his exposition of dream analysis (3). The theory assumes a symbolic relationship between the male and female sex organs and particular classes of objects on the basis of a functional or structural similarity between the organ and that which symbolizes it. The penis is thought to be symbolized by elongated, pointed, or penetrating objects, and the vagina by hollow, rounded, or enclosing objects. Although experimental verification of the theory is lacking, it has had rather widespread acceptance in certain areas of psychology and psychoanalysis and has become an integral part of the literature of projective techniques (1, 6).

A recent experimental study designed to test the hypotheses of sexual symbolism was reported by Levy (5). The specific hypotheses tested were (*a*) that *S*s, when asked to pair symbols with male and female given names, would pair hypothesized "male" and "female" symbols with like-sex names more frequently over the total set of matchings than would be expected by chance, and (*b*) that the learning of "like-sex" pairs of names and symbols would be greater, as measured by retention scores, than the learning of "opposite-sex" pairs. The results of the experiment failed to support either hypothesis.

There are several factors in Levy's design, however, which confound the variables being studied and result in the experiment's lacking conclusiveness. The use of given names as part of the pairing task would seem to permit an unspecified range of earlier learnings to influence *S*s' responses unless names are first equated for all relevant stimulus attributes. A further source of limitation was the use of fifth grade children as *S*s, most of whom are presumably prepubescent. It would appear that the degree of sexual maturity might be a factor relevant to the formation of sexual-symbolic responses.

The purpose of the present experiment was to reassess the hypothesis of sexual symbolism, utilizing as stimuli figures similar to those presented in Levy's study. The primary hypothesis being tested may be stated specifically as follows: When asked to designate stimulus objects as male or female, *S*s respond to elongated, pointed, or penetrating objects as male, and to hollow, rounded, or enclosing objects as female.

A second hypothesis, concerning the differential response of groups of varying degrees of personality integration, is proposed. Since the occurrence of symbolic responses is customarily regarded as a function of socially acquired inhibitions attached to more directly sexual responses, one would expect severe personality disorder, with its disorganization of social-inhibitory processes, to be accompanied by a less frequent occurrence of sexual-symbolic responses than is the case with comparatively well-integrated individuals. Thus, the second hypothesis of this study holds that hospitalized psychiatric *S*s respond to elongated, pointed, or penetrating objects as male, and to hollow, rounded, or enclosing objects as female less frequently than nonpsychiatric *S*s.

METHOD

Subjects. There were two groups of 20 *S*s each. The nonpsychiatric group consisted of ten male and ten female students in an introductory college psychology course. The psychiatric group consisted of ten male and ten female patients at a state mental hospital. Psychiatric and nonpsychiatric *S*s were selected on the basis of age and education in an effort to equate the two groups as closely as possible on those variables. The mean ages for the nonpsychiatric and psychiatric groups were 22.0 years and 27.4 years respectively. The mean educational levels (highest grade completed) were 12.8 and 12.5. For the psychiatric group, patients grossly out of contact, with organic involvement or mental deficiency, and those having received insulin or electroshock treatment within the past six weeks were excluded as *S*s. Schizophrenic disorders were predominant.

[1] The author wishes to express his appreciation to Dr. C. L. Williams, Superintendent of Central State Hospital, for permission to utilize patients as research subjects and to Dr. Morton Wiener, Chief Psychologist, for his advice and encouragement in the execution of this study. The author is indebted to James Norton of the Statististical Laboratory, Purdue University, for statistical help and advice.

[2] Now at The University of Rochester.

Materials. The stimuli in this experiment were the ten figures of hypothesized sexual symbolism used by Levy. Five are elongated, pointed, or penetrating ("male") and five are hollow, rounded, or enclosing ("female"). The figures are essentially "abstract" or without specific object identity; they are variations of circles, pyramids, rods, cubes, etc. In the present study, the figures were enlarged and reproduced in black ink on individual white cards.

Procedure. The *S*s were seen individually by the author. The experimental procedure was third or fourth in a series of other experimental procedures, so there was ample time for the establishment of optimal rapport. The *S*s were presented the following instructions verbally:

> You know, we often think of things other than human beings as having personality. Very often these will be objects around the house or other things that are familiar to us. Some things seem to have a masculine kind of personality and some seem to have more a feminine personality. I have a set of cards here which have designs on them. I'm going to show them to you one at a time and you tell me which they seem to have more—a feminine or masculine personality. Just use your imagination and tell me your first impression. Don't try to think about it too much. Give the very first answer that comes to mind.

Following the instructions, *E* presented the ten figures in an order randomized for each *S*. The *S*s were strongly urged to respond immediately. The few *S*s who persisted in taking longer than two seconds per response were excluded from the study. This procedure was adopted in order to reduce the degree of conscious "problem-solving" types of responses and to prevent defensive, inhibitory responses. It has been held generally that the requirement of immediate response elicits responses that are less guarded or subject to repressive processes and that are in some way more "basic" to the individual's psychological functioning.

RESULTS

Table 1 summarizes the analysis of variance for the three variables of sex, psychiatric classification, and symbol type ("male" or "female"). Scores were the proportion of "correct" responses out of five for each symbol type—i.e., responses supporting the hypothesis of sexual symbolism. The arc-sine transformation for proportions was employed (7).

Comparison of the actual mean with the arc-sine equivalent of a chance expectancy of 50 per cent by *t* test indicated that *S*s' responses were consistent with the hypothesis of sexual symbolism to a degree significantly different from chance expectancy ($t = 27.51$, $p < .0001$). Significant differences between sexes and between psychiatric classifications were found ($p < .05$ in each case). Differences between symbol types failed to reach significance, as was true of the various interaction effects.

TABLE 1

ANALYSIS OF VARIANCE OF PROPORTIONS OF "CORRECT" RESPONSES FOR EACH SYMBOL TYPE

(Data in the form of arc-sine transformation for proportions)

Source	*df*	MS	*F*
Between sexes	1	655.51	4.54*
Between psychiatric classifications	1	931.61	6.46*
Sex × psychiatric classification	1	2.11	
Between *S*s within sex × psychiatric classification cells	36	144.23	
Between symbol types	1	132.61	
Sex × symbol type	1	13.61	
Psychiatric classification × symbol type	1	285.01	1.65
Sex × psychiatric classification × symbol type	1	74.11	
Symbol type × *S*s within sex × psychiatric classification cells	36	173.64	

* Significant at the .05 level.

TABLE 2

ANALYSIS OF VARIANCE OF PROPORTIONS OF *S*S GIVING "CORRECT" RESPONSE TO EACH SYMBOL

(Data in the form of arc-sine transformation for proportions)

Source	*df*	MS	*F*
Between sexes	1	7129	5.85*
Between psychiatric classifications	1	9486	8.08*
Sex × psychiatric classification	1	360	
Between symbols	9	5158	4.24*
Sex × symbol	9	1111	
Psychiatric classification × symbol	9	1828	1.50
Sex × psychiatric classification × symbol	9	1218	

* Significant at the .05 level.

In order to assess between-symbols variation, an additional analysis of variance was performed and is summarized in Table 2. Data were in the form of the arc-sine transformation of the proportion out of ten *S*s who gave the "correct" response to each symbol. The sex × psychiatric classification × symbol interaction was used as the error variance on the assumption that no interaction of this order really occurs. As before, the analysis shows significant differences for the main effects of sex and psychiatric classification. Significant variation between symbols is also shown ($p < .05$). Comparison of pairs of symbols was made by the method of Duncan and Bonner (2). In this analysis, one of the "female" symbols was shown to elicit "correct" responses to a signifi-

cantly lesser degree than all others. None of the other symbols showed significant differences from any other. Thus, the significant variation among symbols is accounted for by this particular one. It is interesting to note that the symbol in question, which might be described as a simple "box" or "cube" is the only one of the ten that is categorized as a sex-symbol on the basis of functional rather than structural similarity to the appropriate sex organ. All other symbols may be regarded essentially as instances of primary stimulus generalization.

Although differences between symbol types and also the symbol type × psychiatric classification interaction failed to show significance in the analysis of variance, a comparison by t test of the mean scores for the psychiatric classifications for each symbol type revealed significantly higher scores for the nonpsychiatric group on "male" symbols than for the psychiatric group. The corresponding difference for "female" symbols was not found to be significant. The presence of the "anomalous" symbol noted above may be responsible for this outcome, however, for when recomputed omitting data from that symbol, $t = 1.92$ with $p < .07$.

DISCUSSION

The results of this experiment strongly support the first hypothesis, that *S*s designate elongated, pointed, or penetrating objects as male, and hollow, rounded, or enclosing objects as female. The results are interpreted as support for the Freudian theory of sexual symbolism.

While the design of the experiment would seem to make plausible a moderate degree of generalization to the subject population, the degree of generalization to other stimulus situations is less clear. It would seem on safe and customary theoretical ground to assume that "response on the basis of sexual symbolism" in this experiment must have occupied some position in the *S*s' response hierarchies prior to the experiment; otherwise, such responses could not have occurred. However, the general dominance of such responses in the *S*s' response hierarchies is not established by this study. In other words, sexual symbolism may be regarded as having been demonstrated in the specific stimulus situation of this experiment and would be expected to generalize to all stimulus situations embodying a significant degree of commonality with the experimental situation. The properties of the generalization gradient, however, remain unexplored. For this reason, application of the results of this study to more complex (e.g., clinical) situations is greatly restricted.

The second hypothesis, that psychiatric *S*s respond to elongated, pointed, or penetrating objects as male, and to hollow, rounded, or enclosing objects as female less frequently than nonpsychiatric *S*s, was also supported. It appears that severe impairment of personality integration is accompanied by a disorganization of symbolic processes. As noted before, a significant difference was shown by t test between the psychiatric and nonpsychiatric groups in their response to male symbols. A similar difference was not found for female symbols. This finding was most likely due simply to the operation of the one "anomalous" female symbol. Thus, one can only report the result of significant differences between psychiatric and nonpsychiatric groups.

The finding of significant differences between sexes, with males responding more frequently on the basis of symbolism, seems to suggest that males are in some way more sensitized to sexual material and that they may be subject to less defensiveness than females in situations with sexual cues. Such hypotheses are consistent with the formulation advanced by Kinsey and his colleagues that ". . . males are more often conditioned by their sexual experience, and by greater variety of associated factors, than females" (4, p. 649). They report specific data indicating that males are more readily aroused sexually by a variety of visual stimuli than are females.

Some areas for subsequent research are suggested. The present study offers support only for sexual symbolism of a primary stimulus generalization type. Symbolism predicated on functional rather than structural similarity between the symbol and the organ symbolized remains unverified. A second area of needed research is the investigation of various subject variables, including period of sexual maturity and physiologic characteristics of sexual functioning, cultural determinants, etc.

SUMMARY

The present experiment attempted to investigate the hypothesis derived from the psy-

choanalytic theory of sexual symbolism that *S*s, when asked to designate objects as male or female, respond to elongated, pointed, or penetrating objects as male, and to hollow, rounded, or enclosing objects as female. The experimental design involved the presentation of ten abstract figures of hypothesized sexual symbolism to psychiatric and nonpsychiatric *S*s of both sexes. The figures, five "male" and five "female," were adapted from those employed in an earlier study by Levy, which failed to support the theory of sexual symbolism.

The total *S*s responded consistently with the hypothesis to a degree significantly greater than chance expectancy, thus supporting the theory of sexual symbolism. It was noted that, while the theory was supported, the extent of generalization of the results of this experiment to other stimulus situations remains unspecified. Males were found to respond consistently with the hypothesis to a significantly greater extent than females, and nonpsychiatric *S*s to a significantly greater extent than psychiatric *S*s. None of the interaction effects proved significant. Further research regarding symbolism predicated on functional rather than structural similarity of the symbol to the organ symbolized is particularly indicated, as the symbols of this study, with one exception, were of the structural type.

REFERENCES

1. BUCK, J. N. The HTP technique: a qualitative and quantitative scoring manual. Part one. *J. clin. Psychol.*, 1948, **14**, 317–396.
2. DUNCAN, D. B., & BONNER, R. G. Simultaneous confidence intervals derived from multiple range and multiple F tests. Report presented to Amer. Statist. Ass., Montreal, Sept., 1954 (mimeographed).
3. FREUD, S. *The basic writings of Sigmund Freud.* New York: Modern Library, 1938.
4. KINSEY, A. C., POMEROY, W. B., MARTIN, C. E., & GEBHARD, P. H. *Sexual behavior in the human female.* Philadelphia: Saunders, 1953.
5. LEVY, L. H. Sexual symbolism: a validity study. *J. consult. Psychol.*, 1954, **18**, 43–46.
6. MACHOVER, KAREN. *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1949.
7. WALKER, HELEN, & LEV, J. *Statistical inference.* New York: Holt, 1953.

Received October 13, 1955.

THE EFFECTS OF THREE VARIABLES ON CHILDREN'S CONCEPTS OF PHYSICAL CAUSALITY[1]

MARTIN L. NASS

Educational Clinic, Brooklyn College

IN RECENT years, a number of studies, stimulated by some of the early work of Piaget (7, 9), have been devoted to a consideration of children's concepts of causal relationships. Piaget's "clinical method" of investigation, as he terms it, made use of individual interviews with children. A record was kept of the child's reply to questions regarding the mode of operation of physical phenomena. Seventeen distinct types of causal explanation, each characteristic of one of three different stages in the development of causal thinking, were distinguished. These stages range from finalism and magic, through animism, to logical deduction. Piaget regards the stage of a child's causal thinking as a reflection of his mental maturity and therefore as related to his chronological age. By the age of seven or eight years the child comes to distinguish between self and outer world and is more realistic in his approach to problems. It is at this age that the more rational types of causal explanation begin to appear.

Piaget's work was extended by other investigators to a consideration of factors other than chronological age on the development of causal thinking. Thus, Oakes (6), Granich (2), and Grigsby (3) considered the differences in causal thinking of children of varying IQ; Deutsche (1) dealt with the effect of socio-economic status; Jones and Arrington (5) with race, and Williams (14) investigated the influence of familiarity with the problem situation in causal reasoning. For a critical review of the literature in this field up to 1943, the reader is referred to Huang (4).

The study reported in this paper analyzes the influence of three factors on the responses of children to various questions regarding physical causality—personality, experience, and question form. The first and third of these factors have not been considered in previous research in this area; the second has been studied to some extent, but in a manner which differs from the present approach (14). The specific questions which the present study proposed to investigate are the following:

1. Does the personality of the child play a role in his concepts of causal relationships?

2. Does the nature of experience which the child has had with the phenomena about which he is questioned affect his ideas of how they operate?

3. Does the form or wording of questions regarding physical causality have an effect on the child's responses?

Personality. No study using Piaget's technique has dealt systematically or specifically with the influence of personality differences on causal concepts, although several investigators (6, 12) speculate on the significance of this factor. It is plausible to suppose that characteristic differences in basic modes of causal thinking exist between socially and emotionally normal children and children with definite tendencies to withdrawal and social avoidance. Often withdrawn children are likely to perceive the environment as threatening, with the anticipated result that "pre-logical" modes of thought should predominate over naturalistic ones. In the present study questions were asked of both withdrawn and emotionally normal children to test the hypothesis that the causal thinking of withdrawn children is at a significantly less mature level than the causal thinking of normal children.

Experience. Systematic study of the child's degree of familiarity with the phenomenon he is questioned about has not been attempted. Studies that discuss this problem (2, 3, 6, 12) have raised certain questions regarding the definition of the term "familiarity." Ordinarily, the amount of experience (or frequency of encounter) the child has had with a phenomenon has been taken to repre-

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, Graduate School of Arts and Science, New York University. A more detailed report is presented in the original manuscript which is also available from University Microfilms, Publication No. 11, 946, $1.35. The author wishes to express his appreciation to Professor Elsa E. Robinson, chairman of the committee, for her guidance throughout the course of this study.

sent his familiarity with it. However, results have shown that items concerning events of weather and movements of celestial bodies produce the more primitive types of causal thinking. The present study has adopted a different criterion for "familiarity" in order to circumvent this difficulty. "Familiarity" as considered herein refers to the child's possible experience with the actual agency of causality of the phenomenon.

A phenomenon can be classified into one of two experience levels, as follows: (*a*), Questions dealing with phenomena in respect to which the child himself can have had, and is likely to have had, direct experience with the agency or mechanics of causation. (*b*), Questions dealing with phenomena with which the child cannot have experienced the causal process directly. By shifting emphasis from the phenomenon itself to the causal processes underlying it, one can readily make this differentiation. It is then possible to test the hypothesis that questions concerning phenomena, the causal agents of which are less accessible to direct experience, yield more nonnaturalistic responses than questions concerning phenomena whose activators are more accessible to the child. In the present study, phenomena of weather and celestial bodies were chosen for the more "remote" experience level since the causal processes underlying them are outside the child's experience while the phenomena themselves are familiar and meaningful to the child. Furthermore, one would expect that the less logical explanations would be given to such items on the basis of the responses to similar questions in previous studies (1, 6).

Question form. Oakes (6) has suggested that the manner in which the question is worded can influence the type of response obtained. Piaget (8) also was aware of the influence of the verbal form of the question. He indicates that a wording form such as, "What makes the sun move?" may suggest the operation of naturalistic forces, and may mask any spontaneous tendencies toward animistic thinking. In Piaget's earlier studies on language (7) he points out that the child's early "Why" questions seem more affective than intellectual in character, and that they generally derive from a disappointment produced by the absence of a desired person or object, rather than signifying verbal curiosity.

A large proportion of the questions asked in previous studies including some of those of Piaget (8, pp. 291, 324), have begun with the word "Why." It is quite possible that such questions direct the child's thinking toward seeking some sort of purpose or motivation. It is one of the purposes of the present study to test the hypothesis that questions worded so as to suggest the possible operation of animistic or dynamic forces yield more nonnaturalistic responses than those which are not so worded.

METHOD

The subjects of the study were 120 children, aged 8-0 to 10-0, enrolled in Grades III and IV of the New York City public schools. Sixty of the children were "emotionally disturbed" and sixty of them were of "normal" emotional adjustment. The disturbed groups consisted of children whose initial referral problems involved "withdrawal." They were selected on the basis of records of the Bureau of Child Guidance of the New York City Board of Education. Normal children were selected on the basis of their over-all school adjustment as noted on the school record and health cards and as judged by their classroom teachers. No child who had been known to a social agency because of adjustment problems was included in the normal group.

The normal and disturbed groups were selected and matched with regard to age, sex, and intelligence test scores. Wherever possible, the technique of matching entailed the selection for the normal group of a classmate of a withdrawn child. All children were of average intelligence (IQ 90–110) as determined by the Pintner-Cunningham Intelligence Test administered in the first grade.

Each of the major groups was divided into two equal subgroups, similar in composition with regard to age, sex, and intelligence-test scores. The experimental design and the composition of the groups with respect to the matching variables are given in Table 1.

Each child was seen individually by the investigator and, after being put at ease in the situation, was given the following instructions:

We are going to play a game. I am going to ask

TABLE 1

COMPARISONS OF THE GROUPS WITH REGARD TO MATCHING VARIABLES

Variable	Group I	Group II	Group III	Group IV
Wording form	A	B	A	B
Adjustment	Normal	Normal	With-drawn	With-drawn
Experience level	1, 2	1, 2	1, 2	1, 2
Mean age	9.2	9.0	9.0	9.0
SD (age)	0.9	0.7	0.7	0.7
Mean IQ	100.5	99.9	100.6	98.4
SD (IQ)	9.2	8.5	9.5	9.5

Note.—Each group is composed of 15 boys and 15 girls.

TABLE 2

LIST OF QUESTIONS USED IN THE STUDY

Experience Level	Wording Form A	Wording Form B	
I	(Why does)	(How does)	the radiator get hot?
	(Why do)	(How is it that)	boats stay on top of the water?
	(Why does)	(How is it that)	a clock ticks?
	(Why does)	(How does)	a whistle blow?
	(Why does)	(How does)	a car move?
	(Why do)	(How come)	leaves fall off the trees?
	(Why do)	(How is it that)	balloons go up in the air?
	(Why do we get shadows?)	(How do shadows come about?)	
	(Why do we have waves?)	(How do waves come about?)	
II	(Why do)	(How do)	we get thunder?
	(Why do)	(How is it that)	we have wind?
	(Why do)	(How is it that)	the stars shine?
	(Why do)	(How is it that)	clouds move?
	(Why do)	(How is it that)	we sometimes have rainbows after the rain?
	(Why does)	(How come)	the sun shine(s) on some days but not on others?
	(Why does night come?)	(How is it that we get night?)	
	(Why is the sky blue?)	(How does the sky get blue?)	
	(Why do we have snow?)	(How does snow come about?)	

you some questions and we will see how many you can answer. Some of these questions are pretty hard, so don't worry if you can't get many. This has nothing to do with your school work and your teacher won't know how you do. I'm just trying to find out what boys and girls know about different things.

The child was then asked the questions in either Form A or Form B of the questionnaire. The materials covered in both forms were selected largely from previous studies and were identical except for the fact that the questions of Form A were posed as "Why's" and therefore might be more suggestive of the operation of dynamic or animistic forces than the questions of Form B, each of which started with "How." The questions are listed in Table 2. Questions from the two experience levels were presented in mixed order. Responses were recorded verbatim, and ambiguous responses were probed by such neutral questions as, "What do you mean?" or "How does that happen?"

Each response was scored independently by three judges.[2] The materials were arranged so as to prevent identification by the judge of either the wording form of the question or the adjustment class of the subject. Responses were classified according to the following scheme:

1. *Nonnaturalistic.* This category included all of Piaget's "precausal" stages which are nonmaterialistic, i.e., responses containing causal statements relating to spirits, magic, or other nonmaterialistic phenomena. It included his motivational, finalistic, moral, magical, animistic, and dynamic classes. Previous studies have shown that areas of overlap among Piaget's 17 categories are so great that it is often virtually impossible to classify a given response definitively (Cf. 4, p. 73).

2. *Phenomenistic.* This is also a "precausal" type, but may or may not be of a materialistic nature. Phenomenistic causality is described by Piaget as follows:

> Two facts given together in perception, and such that no relation subsists between them except that of contiguity in time and space, are regarded as being connected by a relation of causality (9, p. 259).

3. *Naturalistic.* This category included responses which are essentially naturalistic and materialistic. This included Piaget's mechanical and logical categories. Responses need not necessarily be correct to be placed in this category, but must contain underlying ideas which are naturalistic.

[2] The judges were Professor Elsa E. Robinson, Mr. Libero Arcieri, and the author.

RESULTS

Agreement among the judges on categorization of responses was unanimous in 95.7 per cent of the responses. In 3.9 per cent of the responses, agreement in two out of three judgments was obtained. The remaining responses were excluded from analysis. Listed below are examples of responses and their scoring.

Question 1. "Why do leaves fall off the trees?"

a. "Because when winter's coming they couldn't stay on. If they stayed on the snow wouldn't look nice. If they stayed on, winter

TABLE 3
FREQUENCY OF NONNATURALISTIC RESPONSES FOR COMBINED GROUPS

Basis of Grouping	Percentage of Nonnaturalistic Responses	T	p
Normal vs. withdrawn children	18.9 vs. 45.6	2543	$<.001$
Familiar vs. remote causal agency	19.3 vs. 45.7	126	$<.001$
"How" wording vs. "why" wording	15.3 vs. 49.3	2293	$<.001$

Note.—While the T tests were based on rankings of the number of responses in the nonnaturalistic category, percentages are presented here for more meaningful comparison.

TABLE 4
FREQUENCY OF NONNATURALISTIC RESPONSES FOR SUBGROUPS

	Percentage of Nonnaturalistic Responses	
	Normal	Withdrawn
Causal agency "familiar"		
"Why" wording	14.8	54.9
"How" wording	0.4	7.4
Causal agency "remote"		
"Why" wording	48.8	78.5
"How" wording	12.2	43.3

Note.—By Wilcoxon T tests, all four relevant comparisons for each basis of grouping (normal vs. withdrawn, familiar vs. remote, "why" wording vs. "how" wording) are significant at the .001 level.

couldn't come." Scored "nonnaturalistic" by all three judges.

b. "Cause it's cold. They just fall off because it's cold. Because it's winter." Scored "phenomenistic" by all three judges.

c. "They turn different colors and dry up and fall off because they die." Scored "naturalistic" by all three judges.

Question 8. "How is it that we sometimes have rainbows after the rain?"

a. "It's a sign to tell you that the rain is over." Scored "nonnaturalistic" by all three judges.

b. "The sky makes it. Rainbows are in the sky, so the sky makes it." Scored "phenomenistic" by all three judges.

c. "The sun is red and the sky is blue and the clouds are white, so they all come together and you see all the colors together and that's a rainbow." Scored "naturalistic" by all three judges.

Since the "phenomenistic" category accounted for only six per cent of the total responses, analysis was confined to the number of responses falling into the "nonnaturalistic" category, to which the "naturalistic" category was thus essentially the complement. The comparisons were made using Wilcoxon's T tests (13) since the distributions of responses differed significantly from normality and from each other. The results of the comparisons are presented in Tables 3 and 4.

The effects of each factor are clearly shown when the responses of a given group under different conditions are compared. Thus, normal children responding to materials of a more familiar nature which are presented in a nonsuggestive manner give on the average 0.4 per cent nonnaturalistic responses. With less familiar materials this value rises to 12.2 per cent. However, when suggestive wording is introduced the proportions are 14.8 per cent and 48.8 per cent, respectively. Similar differences at a higher level of frequency are noted in the withdrawn groups.

Thus, the causal responses of children to verbal questions concerning physical causality were significantly influenced, as hypothesized, by (*a*) the personality differences tested in this investigation, (*b*) the contact which they may have had with the agency of causation of the phenomena, and (*c*) the wording of the question.

DISCUSSION

Although Piaget does not specifically mention the possible influence of personality factors on level of causal thinking, some of his writings discussing the development of thought processes in children help to clarify his views regarding this variable. Piaget relates the development of more mature thinking to the child's ability to differentiate his subjective world from the external world. During the early stages of development the world and the self are one. Gradually the realms become more distinct, conceptually, until in the final stage the child comes to distinguish between the subjective and objective points of view. This decrease in egocentrism is related to decrease in animism, in anthropomorphic finalism, and in feelings of participation in

external phenomena. The child begins to conceive of reality as constituted by what is common to all points of view taken together (9). The results of the present study with regard to personality differences may be related to this position. The withdrawn child, who feels himself threatened by the environment, may be conceived in Piaget's terms as being hampered in developing the objective point of view, i.e., as more likely to respond syncretically where the emotionally healthier child clearly discriminates self from nonself.

The present findings concerning personality differences leave open the question as to whether the causal factor lies in specific withdrawal tendencies or general maladjustment. Since the "problem" group employed in the present study was restricted to children who emphasized one characteristic mode of handling the environment, i.e., withdrawal, it is not possible to state how a group of aggressive children would respond, or what the results of a comparison between a combined group of "disturbed" vs. "normal" children would be. There are grounds for expecting the withdrawn group, whose retreat is accompanied by fantasy and autism, to produce more nonnaturalistic types of responses than other maladjusted children, but it is not possible to test this hypothesis through the present design.

Experience with a phenomenon as a potential influence on the child's causal concepts is not directly mentioned in Piaget's writings. However, in his later writings on intelligence (10, 11) Piaget extensively discusses the importance of early experiences on later responses. It is not likely that Piaget would deny the importance of experience in determining the level of causal thinking while ascribing a major role to it in the genesis of intelligence. The fact remains, however, that in the volume he devotes to the subject of causal thinking in children (9), Piaget considers neither experimentally nor interpretatively the role of experience in the results which he reports. While the present results uphold the hypothesis that the level of causal thinking may be influenced by experience with the underlying causal agent of a phenomenon, it should be noted that the questions included in the level of more remote experience dealt only with meteorological phenomena. There is a possibility that the differentiating factor may have been something peculiar to these phenomena themselves, rather than experience with their causal processes.

The possible influence of suggestive wording on the responses of children to questions which deal with causality was briefly considered by Piaget, although his exposition of the subject seems to have been logically rather than empirically derived. The present results can be interpreted as lending support to his speculations in this area. The fact that a subsidiary test[3] showed the effects of suggestive wording to be more pronounced with remote than with familiar materials for normal children tends to support some of Piaget's later formulations. He says, "It is true that the more active experience is, the more the reality on which it bears becomes independent of the self and consequently 'objective' (11, p. 367)." Certainly, the more familiar items have been experienced more actively and hence the findings with regard to experience can be reconciled with Piaget's thinking on the subject.

Although these accommodations to Piaget's theory can be made, it should be pointed out that Piaget emphasizes almost exclusively the relation of level of causal thinking to chronological age. On the basis of his exposition, one would expect that most children between the ages of eight and ten years would offer mainly naturalistic causal explanations. The present results, however, show that for some questions the explanations range through most of Piaget's 17 stages at an age at which he contended most should occur near the culminating stages, and fail to support the view, strongly suggested in certain of Piaget's writings, that each developmental level is characterized by a peculiar conception of causality which is generally applied to all the phenomena which the child experiences.

Nevertheless, the present findings at the same time uphold Piaget's insistence that much of the causal thinking of children is prelogical. Although he contended that the naturalistic and logical forms of thinking began to appear by seven or eight years in

[3] A comparison of the differences in number of nonnaturalistic responses between remote and familiar materials showed a significantly greater number for suggestively worded questions than for nonsuggestively worded questions in normal children ($T = 702.5$; $p < .001$).

his sample, and became prevalent in the explanations of the 10- and 11-year-olds, the present study has shown that for the objects and phenomena involved, many of the eight- to ten-year-old children entertained causal ideas which are nonlogical and nonnaturalistic some of the time, and some of the children entertained such ideas most of the time. The data indicate that the nature of children's causal thinking is affected by traits of personality, by the manner in which the question is worded, and by the subject matter of the question.

SUMMARY

This study investigated the effects of three variables on children's concepts of physical causality. It was found that:

1. The nature of the causal thinking of withdrawn children is at a significantly less mature level than the causal thinking of normal children.

2. Questions about phenomena whose causal agents are not accessible to direct experience yielded significantly more nonnaturalistic responses than did questions about phenomena whose causal agents are more accessible.

3. Questions worded so as to suggest the possible operation of "animistic," "supernatural," or "dynamic" forces yielded more such nonnaturalistic types of responses than questions less suggestively worded.

REFERENCES

1. DEUTSCHE, JEAN M. *The development of children's concepts of causal relations.* Minneapolis: Univer. of Minnesota Press, 1937.
2. GRANICH, L. A qualitative analysis of concepts in mentally deficient schoolboys. *Arch. Psychol., N. Y.*, 1940, No. 251.
3. GRIGSBY, OLIVE J. An experimental study of the development of concepts of relationships in preschool children as evidenced by their expressive ability. *J. exp. Educ.*, 1932, **1,** 144–162.
4. HUANG, I. Children's conception of physical causality: a critical summary. *J. genet. Psychol.*, 1943, **63,** 71–121.
5. JONES, F. N., & ARRINGTON, M. G. The explanations of physical phenomena given by white and Negro children. *Comp. Psychol. Monogr.*, 1945, **18,** No. 5.
6. OAKES, M. E. Children's explanations of natural phenomena. *Teach. Coll. Contr. Educ.*, 1947, No. 926.
7. PIAGET, J. *The language and thought of the child.* New York: Harcourt, Brace, 1926.
8. PIAGET, J. *The child's conception of the world.* New York: Harcourt, Brace, 1929.
9. PIAGET, J. *The child's conception of physical causality.* New York: Harcourt, Brace, 1930.
10. PIAGET, J. *The psychology of intelligence.* New York: Harcourt, Brace, 1950.
11. PIAGET, J. *The origins of intelligence in children.* New York: International Universities Press, 1952.
12. SARVIS, B. C. *A clinical study of the developmental aspects of child thought.* Unpublished doctor's dissertation, Univer. of Kansas, 1939.
13. WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Co., 1949.
14. WILLIAMS, M. *The influence of specific experiences on children's concepts of causal relations.* Unpublished doctor's dissertation, Univer. of Michigan, 1940.

Received November 3, 1955.

SPEED OF FORM RECOGNITION AS A FUNCTION OF STIMULUS FACTORS AND TEST ANXIETY[1]

FREDERICK C. ROCKETT[2]
The Dow Chemical Company

DURING the past decade, studies have appeared in the psychological literature reflecting an increasing interest in the etiology as well as the manifestations of color shock, generally considered a sign of neuroticism, in the Rorschach test (18). The results of many studies suggest that certain inkblots seem to provoke "color-shock" responses whether the blots are colored or achromatic (1, 2, 6, 11, 12, 16, 17, 19). All of these studies, however, suffer from the lack of a proper control group representing neurotic or highly anxious *Ss*. Studies employing a proper control group have been more successful in supporting the hypothesized relationship between color, neuroticism, and certain responses to colored inkblots (3, 4, 13), but none has found more than tenuous evidence.

However, from the speculations of various students of color shock have come several conceptions regarding its etiology. Schachtel (21) postulates an innate responsiveness of neurotics to color, while Wallen (24) suggests that positive and negative affective responses to color are due to the association of color with reward and punishment respectively. Siipola (22) believes that hue-form incongruity or the complexity of the stimulus may greatly alter response content. A number of studies suggest that stimulus complexity is the main determinant of color-shock indices (5, 11, 17). In general, there are two approaches to the interpretation of color shock. One is through characteristics of subjects, and the other through stimulus characteristics.

The present study is an attempt to test various color-shock hypotheses and their underlying theoretical assumptions through an analysis of response time. The color-affect theory (Schachtel, Wallen) concerning color shock in the Rorschach test suggests that highly anxious individuals have a relatively delayed reaction to brightly colored cards because of a negative affective response to color. Nonanxious individuals are comparatively unaffected by color and consequently have faster reaction times. The following hypotheses were framed for investigation:

Hypothesis 1. Highly anxious individuals take significantly longer to recognize colored ambiguous figures than less anxious persons.

Hypothesis 2. Recognition times for figures having disliked colors are significantly longer among highly anxious persons than for figures having pleasant colors.

Hypothesis 3. Disliked colors are more often linked with unpleasant concepts (e.g., blood) in highly anxious individuals than in less anxious ones.

Students of color shock who emphasize the role of stimulus complexity suggest that the relationships proposed in Hypotheses 1, 2, and 3 are less important than such factors as figure-ground contrast through the juxtaposition of various colors, the intrinsic difficulty of the figure, and other stimulus-complexity determinants.

Hypothesis 4. Stimulus attributes contributing to complexity contribute significantly more variance to total performance than anxiety level or color preference.

Cox and Sarason (4), like others (7, 23), postulate that anxiety has drive characteristics, while Miller (14) suggests that any strong stimulus may have drive characteristics. From Hullian behavior theory (10), it follows that with increasing drive strength, response time to simple stimuli becomes faster and suprathreshold responses more frequent and stronger in a *free response* situation. Thus, the degree of response competition, the number and strength of simultaneously elicited responses, is positively related to drive level and stimulus complexity. When response competition is strong, however, then response time should be faster for less anxious than for more anxious *Ss* in a situation in which only *one response* can be made. Godbeer (8) has shown that in such instances multiple approach-avoidance conflicts occur. It follows that re-

[1] A condensation of a dissertation presented in partial fulfillment of the requirements for the Ph.D. degree, University of Texas, 1955.

[2] The author is grateful to the faculty of Wichita University for their hospitality and aid in the completion of this study. Dr. N. H. Pronko deserves particular recognition.

sponse time should be positively related to the number and strength of these conflicts.

Hypothesis 5. Factors influencing drive level (anxiety, color preference) are negatively related to response time when stimulus complexity is low, but positively related to response time when complexity is high and only one response is allowed.

METHOD

Experimental design and materials. The hypotheses were tested in a single design in which a test series followed a warm-up series of four forms from the Holtzman Form-Recognition Test (9).

The Holtzman Forms are line drawings of common animals, birds, reptiles, and humans in which identification is delayed by omission of parts of the figure and by the addition of superfluous lines. The average response time to the figures selected for this study ranges from 1.10 seconds to 3.00 seconds. Correct identifications took place 95 per cent of the time in the original standardization with college students. The Holtzman Forms were used in order to reduce the difficulty encountered in many Rorschach studies of dealing with stimuli so complex that many responses are commonly given to a card. In such instances, it is often difficult to determine what is or is not a deviant response, and statistical problems of analysis are greatly complicated.

Warm-up cards were presented so that novelty would not delay responses, thereby contributing to the total variance in response time of the main study. A pilot study and the work of Doidge (5) showed that response times reach an asymptote with the administration of four practice cards. The order of presentation was the same for each subject, progressing from lesser to greater levels of difficulty with the most difficult warm-up figure being somewhat less difficult than the easiest test figure.

Colors used in the test series were selected for closeness to the four psychological primaries and for brightness. The Munsell designations of the colors are presented in Table 1. The patterns in which the colors were presented were selected to represent a wide range of possible patterns, to follow the work of Doidge, and to lend themselves to fairly uniform duplication. The four patterns were wave, speckled, splotched, and solid. Colors and patterns were combined by spray-painting each of the patterns in each of the colors on sixteen glass plates. Such plates, when placed before the Holtzman Figures, gave the appearance of colored patterns and figures on single surfaces. The wave and splotched patterns were stencilled for greater uniformity.

The viewer used was specially constructed, using the main component of the Harvard Tachistoscope as a model. It was 9¼″ square and 21¾″ long. The inside of the viewer was painted flat black on all surfaces except the areas around two recessed light sources, which were glossy white. The light from two long unfrosted bulbs of 25 watts was made diffuse by the interposition of frosted glass.

TABLE 1
MUNSELL COLOR VALUES OF TEST SERIES

Hue	Lightness	Chroma
Greenish-yellow (7.5 Y)	8	10
Red (5.0 R)	4	14
Blue (5.0 B)	4	8
Yellowish-green (2.5 G)	5	8

Of the several scales which purport to measure neuroticism and anxiety, only two have been utilized to such an extent that some confidence can be expressed in their utility. These two scales are the Taylor Manifest Anxiety Scale (23) and the Sarason Anxiety Scale (20). The Sarason was used rather than the Taylor because it appears relatively innocuous. The *Ss* were not given the idea that they were being evaluated critically, and were therefore not forced into a state of anxiety due to an adventitious situational factor.

An arabic-greco-latin square (AGLS) design was employed, composed of three independent latin squares superimposed on each other so that each of three independent factors appears once, and only once, in each row, each column, and in combination with each of the other two factors. Sixteen different AGLSs were necessary to represent all possible combinations of the three experimental stimulus-variables (figure, color, pattern) in the first column of the AGLS (Trial I). Since four *Ss* are needed to complete each square, 64 *Ss* constitute one complete design. The addition of three levels of anxiety as an independent variable required a total of 192 *Ss*. This design allowed for the testing of the interaction of the main variables with anxiety and AGLS, as well as allowing the testing of these variables individually.

The first response to the test series for each *S* was analyzed in a $4 \times 4 \times 4 \times 3$ analysis-of-variance design without replication. It was possible to carry out this analysis while maintaining orthogonality because each variable was presented once in the first trial of each square.

Subjects. The *Ss* were 192 male students attending Wichita University. The entire Air Reserve Officer Training Corps, all freshmen of the Army Reserve Officer Training Corps, and several classes in introductory psychology served as the population from which the sample was drawn.

Procedure. The Sarason Anxiety Scale (19) was administered during regularly scheduled classroom hours. All *Ss* falling in the middle 15 per cent and the two extreme 15 per cent segments of the distribution obtained on the Sarason Scale were eligible for the main study. Samples from the middle as well as the extremes were drawn because a previous study (7) had indicated a marked curvilinear relationship between performance and anxiety level. These men were called by telephone for individual appointments. Only two *Ss* refused to participate.

Having arrived at the experimental room, the *S* was seated before a viewer. The room itself was a 9-ft. × 10-ft. windowless chamber containing only material relevant to the present study. The *S* was then given a color-blindness test using the Dvorine Color Charts. If *S* passed this test, *E* read the following instructions; otherwise the session was terminated at that point. Eleven *Ss* were discarded as color blind.

Look into the viewer. Presently a series of cards

will be shown to you. On each card there will be a figure, such as an animal, bird, insect, or human. It is your task to identify these figures as quickly as possible, using the specific name in each case, such as chicken, lizard, etc. I will be timing you with a stop watch.

Remain in position for the next card after the light goes off.

Are you ready?

While these instructions were being read, *E* prepared the stimulus materials for the first presentation. The first card of the warm-up series was inserted into an especially constructed holder attached to the end of the viewer opposite to the viewing aperture. A clear glass plate was put just in front of the figure to insure uniformity of illumination of the figures for both the warm-up and test series.

Immediately after the instructions were read, the first warm-up figure was presented by turning on the lights in the darkened viewer. When a response was made, the lights were turned off, the response time and response were recorded on a form, and the second stimulus prepared. A shield was placed in the position ordinarily occupied by the stimulus holder during changes of stimuli to prevent *S* from seeing the characteristics of the forthcoming material. The other stimuli followed in the same fashion without further communication with *S*, except to indicate that another figure was about to be presented. Care was taken to give *S* no clue to the shift to colored stimuli after the fourth warm-up card.

A 60-second time limit was allowed for making each response, after which time the light in the viewer was turned off and preparation for the next trial begun. If *S* wanted to give up before the 60 seconds had passed, *E* encouraged him to continue trying.

After the response times and responses were recorded, *S* was asked to stand before a table placed near the viewer. On the table were the four glass plates bearing the color-pattern combinations that *S* had seen during the test series. Under the plates were pieces of white paper of approximately the same reflectance as the photographic paper on which the figures were printed. The color-patterns then had an appearance similar to those which were seen in the viewer except that the figures were absent. At this time the order of preference and first association to the color patterns were recorded. Lastly, questions designed to assess the reliability and validity of the Sarason scale as a measure of neurotic anxiety were asked. They were:

1. The questionnaire that you took for me tried to get at how confident you are in taking tests. Your score on the questionnaire fell in the (15 per cent most confident, 15 per cent least confident, or middle 15 per cent). Do you think that this is reasonable?
2. How confident are you socially?
3. Have you ever had any kind of personal counselling?
4. What kind, and where?
5. How well adjusted do you feel? How well satisfied with yourself as a person are you?
6. To what extent do you think that you would fear a personality evaluation such as taking a battery of tests and being interviewed by a psychiatrist?

RESULTS

Responses to the above questions were categorized in three divisions for statistical purposes. These three divisions represent the extremes and the average of the range of responses. Chi square revealed that responses to questions 1, 2, 5, and 6 are highly related to the Sarason-scale scores ($p < .001$). That is, high anxiety as indicated by the Sarason scale was highly related to responses thought to indicate neuroticism. No difference in frequency of reported attendance to a counseling service was found for the three anxiety groups.

To normalize the typically skewed distribution of response-time scores, an inverse square-root transformation was applied. All statistical calculations involving response-time data utilized transformed scores.

The results of an analysis of variance of response time during warm-up trials showed significant differences between anxiety levels ($p = .01$), trials ($p = .001$), and the interaction between anxiety and trials ($p = .01$). The highly anxious group proved to be somewhat slower in response time than the other two groups, although they registered the fastest mean response time on the first trial. In each trial thereafter in the warm-up series, the high anxiety group became progressively slower relative to the medium and low anxiety groups.

Since figures were presented in the same order for all *S*s, trials and figures are confounded. Nevertheless, because the warm-up figures were presented in an ascending order of difficulty according to standardization information, and because the median response time for each trial or figure became progressively shorter except for the last of the series, it is clear that there were strong practice effects which obscured the response-time characteristics of the figures.

The comparison of the mean difference score for the last warm-up and the first of the colored cards for the high and low anxiety groups, the two with the greatest mean difference, showed no significant difference. All groups displayed similar increases in response time, significant beyond the .001 level of confidence, to the first colored card.

Analysis of variance of response times to the first colored card revealed the powerful role played by the color ($p < .001$), figure

TABLE 2
ANALYSIS OF VARIANCE FOR THE ARABIC-GRECO-LATIN SQUARE DESIGN

Source of Variation	*df*	Mean Square	*F*	*F*†
Independent Observations:				
Different AGLS	15	63.0	1.60	
Anxiety levels	2	75.0	1.90	
Anxiety × AGLS	30	36.0	.93	
Residual between *Ss*	144	39.0		
Total between *Ss*	191			
Correlated Observations:				
Trials	3	47.0	3.00*	1.6
Color	3	380.0	24.50***	
Pattern	3	966.0	62.30***	
Figure	3	632.0	40.80***	25.5***
Trials × AGLS	45	29.0	1.90**	
Color × AGLS	45	20.0	1.30	
Pattern × AGLS	45	19.0	1.20	
Figure × AGLS	45	25.0	1.60*	
Trial × Anxiety	6	20.0	1.30	
Color × Anxiety	6	12.0	.77	
Pattern × Anxiety	6	7.0	.45	
Figure × Anxiety	6	12.0	.77	
Residual	360	15.5		
Total within *S*	576			
Total for experiment	767			

* $p < .05$, ** $p < .01$, *** $p < .001$.

† These additional *F* ratios have been computed using the trial AGLS interaction and the figure AGLS interaction to test the significance of trials and figures, respectively.

($p < .001$), and color-pattern ($p < .001$) variables. Anxiety level did not influence response time to an appreciable degree. The interactions of pattern and figure ($p = .01$) and of color, pattern, and figure ($p = .05$) also played a significant role in response-time determination.

Table 2 summarizes the results of the analysis of variance for the AGLS design. Colors, patterns, and figures were found to be potent factors in the determination of response time while trials and the interactions of both trials × AGLS and figures × AGLS were significant but lesser contributors. Anxiety level did not play a significant role in response-time determination in the test series, either alone or in combination with other variables.

Analysis of color, pattern, and color-pattern preferences showed no difference among anxiety groups. Colors have a wider range of preference frequencies than do the patterns, indicating a greater range of affectivity for color. Chi-square analysis of the frequency of each degree of preference for each color shows that colors tend to have significantly different ($p = .01$) frequencies at each degree of preference. Differences in frequency of preference for patterns were significant ($p = .01$) only for the most preferred category. Ranking colors in order of response time and comparing the order to that found for color preference indicated that the relationship between response time and preference is the exact reverse of the prediction from color-affect theory that negative affect to color is associated with longer response times.

Further analysis of response time for the three anxiety groups and for degree of color preference uncovered no significant relationship between anxiety group, color preference, and response time. Combining the anxiety groups did not alter the finding of no significant difference in response time between degrees of color preference.

Only a few associations to the color patterns were regarded as unpleasant. Some typical associations were linoleum, flowers, waves, and germs. There were only 5 "blood" associations in all 768 opportunities, which included 48 presentations of a red splotched pattern. This number was too small for statistical consideration in relation to anxiety groups.

In order to sharpen the anxiety variable, an especially selected maladjusted group was compared in several ways with an especially selected well-adjusted group. The *Ss* who had indicated maladjusted responses on any two of the three interview items of social anxiety, fear of psychiatric evaluation, or degree of adjustment, or who had indicated that they had had psychiatric treatment, were grouped and their various scores compared with those of the medium and low anxiety categories. This gave a total of 45 *Ss* in the maladjusted group to be compared with 109 *Ss* who neither met the above criterion nor scored in the highly anxious range on the Sarason Anxiety Scale. No evidence was found to indicate that this further refinement of the anxiety variable yielded results different from those reported.

DISCUSSION

Hypothesis 1, that highly anxious individuals have longer response times than those

less anxious, was not substantiated. The three anxiety levels did not have significantly different mean response times in the test series. Similarly, no differences were found among anxiety levels in response time between the last warm-up card and the first card of the test series.

Hypothesis 2, that disliked colors result in longer response times than liked colors for high anxiety *Ss*, was also unsupported. No significant difference in response time between color preferences for highly anxious *Ss* was found. It should be recognized, however, that several variables of known significance were uncontrolled in this test. Complete control of color, color pattern, and figure was not possible, and it is likely that some imbalance was present, particularly for color.

Strengthening the notion that color preference is not a major factor are the negative results of a pilot study in which color pattern, color preference, and figure were independent variables. These findings are consistent with evidence (14) indicating that color preferences are largely learned and are quite ephemeral, being amenable to rapid change. Further, Warner (24) exhaustively studied the color preferences of various psychiatric groups, finding little of consequence other than the usual male-female differences. It is likely therefore, that most persons are not affected by color preferences to the same degree that they are by other factors.

As an incidental observation and contrary to usual clinical belief the color red was neither disliked by more highly anxious persons than by others, nor was it associated with longer response time.

A weakness in the study is indicated by the lack of affect-laden associations to the color patterns. This defect vitiates considerably the testing of Hypothesis 3. However, it seems interesting, considering common practice, that the association "blood" was so rare an occurrence, despite the frequent use of a red splotched pattern.

Complication theory seems best supported by these results. As Doidge and others have indicated, factors contributing to stimulus complexity delay response time, whereas other factors remain relatively unimportant in this respect. Hypothesis 4 was confirmed by the powerful stimulus effects of colors, color pattern, and figure in response-time determination; while the influences of such respondent variables as color preference and anxiety were negligible. The significant difference between trials and the significant interaction between trials and AGLS cannot be dealt with by complication theory in its present form, but these findings represent order effects, not centrally relevant problems in the relationship of affect to color.

Hypothesis 5 was only partly confirmed because factors which theoretically could have altered drive level did not function consistently. Neither anxiety level nor color preference had any discernible effect on response time in the test series, although anxiety level and the interaction of anxiety and trials were significant determinants of response time in the warm-up series.

Inspection of the means for each trial at each anxiety level for the warm-up series suggested that response time for the highly anxious *Ss* became progressively slower relative to the medium and low anxiety *Ss*. Conceivably, a multiple approach-avoidance conflict may have been greater for high anxiety *Ss*. If this explanation were true, however, one would expect that the greater difficulty of the test series would increase the conflict, further slowing response time. This was not found. Thus, warm-up findings may have been due to chance alone, or it may be that some other explanation is possible. Future research could profitably explore greater ranges of drive level and stimulus complexity and their relationship.

It can be said that little confirmation for any explanation of color shock was found except for complication theory. The idea that response time is largely a matter of stimulus attributes and not the anxiety of *Ss* received strong support. However, because they generate testable hypotheses combining stimulus and subject characteristics, more general theories of behavior seem to offer the most promising route to a fuller understanding of responses to ambiguous figures, including Rorschach responses.

SUMMARY

An attempt was made to evaluate the usefulness of several theories offering explanations of color shock as it is said to occur in Rorschach testing. An arabic-greco-latin square design was utilized in the test series of a study

(N = 192) which investigated the influence of color, pattern of color, figure, and anxiety level on response time. The influence on response time of stimulus attributes affecting complexity was confirmed. The hypotheses that highly anxious persons are greatly affected by colors are not confirmed. Partially verified were predictions made from Hullian behavior theory. It was found that drive level and figure difficulty were related to response time in that highly anxious persons became slower with increasing difficulty of figure during warm-up trials. During the test series, however, the predicted relationship was not found. Further work should explore more fully the extremes of the stimulus complexity and anxiety variables.

REFERENCES

1. ALLEN, R. M. The influence of color in the Rorschach test on reaction time in a normal population. *J. proj. Tech.*, 1951, **15**, 481–485.
2. ALLEN, R. M., MANNE, S. H., & STIFF, M. The role of color on Rorschach's test: a preliminary normative report on a college student population. *J. proj. Tech.*, 1951, **15**, 235–242.
3. BUKER, S. L., & WILLIAMS, M. Color as a determinant of responsiveness to Rorschach cards in schizophrenia. *J. consult. Psychol.*, 1951, **7**, 196–202.
4. COX, F. N., & SARASON, S. B. Test anxiety and Rorschach performance. *J. abnorm. soc. Psychol.*, 1954, **49**, 371–377.
5. DODGE, W. T. Influence of color upon form perception. Unpublished master's thesis, Univer. of Texas, 1953.
6. DUBROVNER, R. J., VON LACKUM, W. J., & JOST, H. A study of the effect of color in productivity and reaction time in the Rorschach test. *J. clin. Psychol.*, 1950, **6**, 331–376.
7. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–126.
8. GODBEER, E. Factors introducing conflict in the choice behavior of children. Unpublished master's thesis, Yale Univer., 1940.
9. HOLTZMAN, W. H. Emotional instability and perceptual blocking induced by color distraction. Paper read at South. Soc. Phil. Psychol., Austin, Texas, April, 1953.
10. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943. Pp. 185–225.
11. LAZARUS, R. S. An experimental analysis of the influence of color on the protocol of the Rorschach test. *J. Pers.*, 1948, **17**, 132–135.
12. MEYER, B. T. An investigation of color shock in the Rorschach test. *J. clin. Psychol.*, 1951, **7**, 367–370.
13. MIALE, FLORENCE R., & HARROWER-ERICKSON, MOLLY R. Personality structure in the psychoneuroses. *Rorschach Res. Exch.*, 1940, **4**, 71–74.
14. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens, (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
15. NORMAN, R. D., & SCOTT, W. A. Color and affect: A review and semantic evaluation. *J. gen. Psychol.*, 1952, **46**, 185–225.
16. PERLMAN, J. A. Color and validity of the Rorschach 8–9–10%. *J. consult. Psychol.*, 1951, **15**, 122–126.
17. RABIN, A. I., & SANDERSON, M. H. An experimental inquiry into some Rorschach procedures. *J. clin. Psychol.*, 1947, **3**, 216–225.
18. RORSCHACH, H. *Psychodiagnostics*. New York: Grune & Stratton, 1942.
19. SAPPENFIELD, B. R., & BUKER, S. L. Validity of the the Rorschach 8–9–10% as an indication of responsiveness to color. *J. consult. Psychol.*, 1949, **13**, 268–271.
20. SARASON, S. B., & GORDON, E. M. The test anxiety questionnaire: scoring norms. *J. abnorm. soc. Psychol.*, 1953, **48**, 447–448.
21. SCHACHTEL, E. G. On color and affect: Contributions to an understanding of Rorschach's Test II. *Psychiatry*, 1943, **6**, 393–407.
22. SIIPOLA, ELSA M. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.
23. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
24. WALLEN, R. The nature of color shock. *J. abnorm. soc. Psychol.*, 1948, **43**, 346–356.
25. WARNER, S. J. The color-preferences of psychiatric groups. *Psychol. Monogr.*, 1949, **63**, No. 6 (Whole No. 301).

Received November 3, 1956.

AVOIDANCE CONDITIONING OF VERBAL BEHAVIOR WITHOUT AWARENESS: A PARADIGM OF REPRESSION[1]

CHARLES W. ERIKSEN[2] AND JAMES L. KUETHE

The Johns Hopkins University

SEVERAL recent experiments have suggested that the principles of learning theory may be profitably applied to the study of language and verbal behavior (2, 4, 6). These studies have shown that the verbal behavior of a subject can be manipulated through the use of differential reinforcement and in many cases the subjects remain unaware that their behavior has been modified, i.e., they are unable to state in what way their behavior has changed or even that it has changed. While these experiments have dealt with overt verbal behavior, it seems reasonable that implicit verbal behavior or thinking might similarly be affected. If so, these experiments have some interesting implications for research on psychological defense mechanisms in that they offer suggestions as to how defenses might be learned and how they might operate.

It is an empirical observation that responses followed by punishment or anxiety tend to be eliminated or replaced by other responses. If thoughts or implicit verbal behavior are considered as responses subject to manipulation by reinforcement, we can expect that thoughts that have become conditioned stimuli for anxiety will be eliminated or replaced by other thoughts. Dollard and Miller (3) have presented a detailed theoretical treatment of defense mechanisms along these lines.

Such a theoretical account of the learning of defense mechanisms would gain support if it could be shown that thoughts or associations that were followed by anxiety were eliminated or replaced. It was our purpose in the present study to determine, first, whether verbal associations that the subject (*S*) had formed during his life experiences could be changed by punishment in the form of electric shocks, and second, whether these changes in verbal associations could be produced without the *S*'s becoming aware that his behavior had been modified. We also wished to determine if such effects would persist when the experimental task was changed.

METHOD

The basic technique used in this study was a modified form of the word association test. Here the same set of stimulus words was administered to *S*s a number of times in an unbroken sequence. Table 1 shows the 15 stimulus words comprising the association list that was used. It was assumed that each word on this list constituted a stimulus having a hierarchy of associations or responses that had been acquired by *S* during his life experiences. From a pilot study we had found that when the same word list is repeated a number of times, *S*s tend to continue to repeat their original associations. We wished to determine whether strong electric shocks administered every time certain selected associations occurred would result in these responses dropping out and being replaced by other associations.

Prior to beginning the experimental sessions with an *S*, the *E* randomly designated five of the stimulus words in the list which were to be the *critical* stimuli for this particular *S*. During the first trial the *S*'s associations to these five words, whatever they happened to be, were punished by a strong electric shock. Without pause the *S* was given a second trial on the 15 stimulus words followed by a third trial and so on for a total of ten trials or until the *S* had reached a criterion of two successive trials without the occurrence of any one of the five punished associations. On each trial the stimulus words occurred in a different sequence. During these succeeding trials on the list, *S* was administered a strong electric shock every time he responded with one of the five associations that had been shocked on the first trial. Shocks were administered only when these "taboo" associations occurred.

Shocks were administered through ankle electrodes attached to the *S*'s left ankle. The source of the shock came from an inductorium and a 3 v.d.c. source. Faradic stimulation was used. The intensity of shocks varied slightly between *S*s due to such factors as differences in skin resistance, but in all cases it was evident from the *S*'s comments and behavior that the shocks were quite painful.

In order to maximize the opportunities for obtaining "learning without awareness," the *S*s were misinformed as to the nature of the experiment and the reason for the electric shocks. They were told that the purpose of the experiment was to determine the maximum speed of mental associations. They were instructed that shocks would occur under two conditions: one, if their association to a stimulus word was too slow for that particular stage of practice; and the second was a condition that the *E* could not divulge to them in advance, but they might discover it on their own during the session and thus be able to avoid getting the shocks.

[1] This research was sponsored by Research Grant M-1037 from the National Institute of Mental Health.

[2] Now at the University of Illinois.

TABLE 1
STIMULUS WORDS

Careless	Lovely
Perfect	Hasty
Cheerful	Gawky
Tidy	Rural
Rotten	Thrifty
Costly	Old
Small	Boastful
Polite	

The *S*s were instructed to respond to each stimulus word with the first word that came to mind and as quickly as possible. To insure that they understood the nature of the task, several practice words were administered. They were also informed that the same list of words would be repeated over and over, but each time the words would occur in a different sequence. During the experimental session *S*'s association time to each stimulus word was measured to the nearest hundredth of a second by means of an electronic voice key. The timer was started by the sound of *E*'s voice and manually stopped by *E* when *S* responded.

At the conclusion of this learning phase of the experiment, *S*s were informed that this experiment was over and were then asked to do one more task before leaving. This consisted of chained associations to the stimulus words. The *S*s were told that the same stimulus words would be repeated again, but this time they were to say all the words that came to mind and continue doing so until the *E* said "stop." Fifteen seconds were allowed for chained associations to each stimulus word. Before starting this phase of the experiment, *S*s were all carefully instructed that there would be no further electric shocks.

In carrying out the chained associations in this experiment, we were interested in two things. First, we wished to see if the punished association would occur when the *S* was required to give a large number of responses, and second, we were interested in determining if the effects of punishment would be generalized to a situation somewhat different from the original learning situation, in which the threat of shocks had been removed.

At the conclusion of the chained associations the *S* was closely interviewed to determine his ability to verbalize the true nature of the experiment and changes in his behavior. All *S*s were asked the following questions: Did you discover how to avoid getting shocks? Did you do anything to keep from getting shocked? What did you do? Did you ever find yourself withholding a response word and trying to think of something else to say? (If a positive answer was obtained, *S* was asked if he could recall any instances where this occurred.) Those *S*s who reported deliberately withholding responses and substituting other responses for them were also questioned as to whether they had to continue doing this throughout the experiment.

In addition to the questioning, *S*s were shown a list of the stimulus words and asked to indicate which words had been followed by shock and what response word they had given when the shocks occurred.

Subjects. Thirty-one male undergraduate students served as *S*s in the experiment. All were obtained from the elementary psychology course at The Johns Hopkins University. As part of the course requirements, they were required to serve as *S*s in psychological experiments.

RESULTS

Awareness. On the basis of the interview data it was possible to divide the *S*s clearly into high- and low-level-of-awareness groups. The high awareness *S*s were able to state the real basis for shocks and to describe what they had done to avoid them. They were also able to identify correctly three or more of the critical stimulus words as well as the response they had given that led to punishment. All reported deliberately withholding punished responses while trying to think of another response to substitute for it. However, they all reported that after a few trials it was no longer necessary to withhold the punished responses deliberately since the new, or safe, response began to occur "automatically."

The low awareness *S*s, on the other hand, were unable to state any reasons for the shock other than the misinformation the *E* had given them prior to the experiment. Occasionally they expressed hypotheses that were quite unrelated to the real basis for shock and which were quite at variance with their actual behavior. They were unable to provide an accurate description of changes in their behavior and all stated that they had not learned how to avoid getting shocks. When asked to indicate which of the stimulus words had been followed by shock, their selection was random. In no case were their selections more than 20 per cent correct.

It was possible to classify eleven of the *S*s as having a high level of awareness or insight and another eleven *S*s as having a low level of insight. There were another five *S*s who could not be clearly placed in either the insight or the noninsight group. These *S*s seemed to have a vague idea as to the reasons for shock and what they needed to do to avoid them, but their verbalizations were not definite enough to permit clear-cut decisions. We have labeled these *S*s as the partial insight group. The remaining four *S*s resisted classification on our insight criteria. Three of these *S*s had adopted an incorrect hypothesis concerning the shocks. While their hypotheses were incorrect, they were nonetheless effective in avoiding shocks. For example, two of these *S*s had the hypoth-

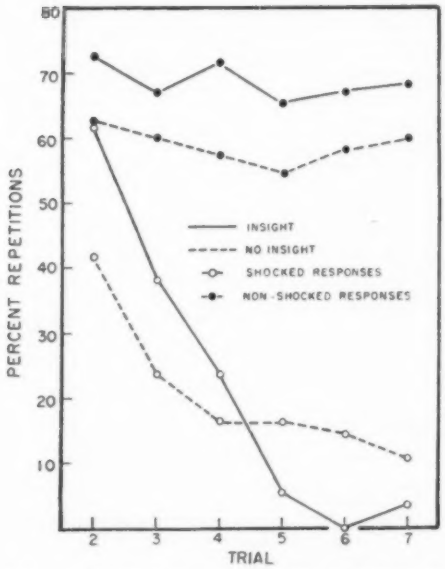

FIG. 1. PERCENTAGE OF FIRST-TRIAL RESPONSES REPEATED ON SUCCCEEDING TRIALS AS A FUNCTION OF PUNISHMENT

esis that shocks occurred when they gave synonyms and they had changed all of their responses to avoid synonyms. The other two *Ss* had resorted to using only two response words, such as, "good" and "bad" for all stimulus words. The data from these four *Ss* were excluded from further analysis.

Since we were successful in obtaining two levels of awareness or insight in our *Ss*, we have treated insight as an independent variable in the remainder of the data analyses. We have not presented the data for the partial insight group, however, due to the small number of *Ss* that fell into this group.[3]

Avoidance learning. Both the insight and noninsight groups were quite successful in learning to avoid the punished associations. As Fig. 1 shows, the number of repetitions of punished responses decreases rapidly for both groups while the number of nonpunished responses that were repeated from trial to trial tends to remain constant. For the insight group, a mean of 68.2 per cent of the first-trial, nonpunished response words were repeated on Trial 7 while only 3.6 per cent of the punished responses were repeated on this trial. A t test of the difference between these two means resulted in a t value of 12.2 which with ten degrees of freedom was significant beyond the .001 level. The same comparison for the noninsight group shows the mean percentage of repetition of the first-trial, nonpunished responses on Trial 7 to be 60.0 per cent as compared with a mean percentage of 10.9 for the punished responses. The t value of 6.21 is again quite significant ($p < .001$).[4]

While both the insight and noninsight groups show marked and reliable evidence of avoidance learning in this experimental situation, there does not appear to be any reliable evidence of a difference in rate of learning between these two groups. From Fig. 1, it appears that the noninsight group showed a tendency for more avoidance of shocked responses on Trials 2 through 4, but a t test based upon the mean percentage of shocked responses for each *S* in the two groups summed through these three trials does not approach statistical reliability ($t = 1.42$, $p > .10$). Also the same comparison between the insight and noninsight *Ss* on Trials 5 through 7 yields a t of 1.44 which is again insignificant ($p > .10$).

From Fig. 1 it appears that the noninsight group tended to give fewer first-trial responses to the noncritical stimuli than did the insight group. To determine whether this difference was reliable, the mean percentage of first-trial responses given by each *S* on Trials 2 through 7 was computed. The means of these mean percentages were 68.8 for the insight group and 58.8 for the noninsight group. The resulting t of 1.22 with 20 df does not approach

[3] In nearly all of the subsequent comparisons it was found that the partial insight *Ss* were intermediate between the insight and the noninsight groups although their performance more closely resembled that of the insight *Ss*.

[4] An over-all analysis of variance of these data was not justified due to the marked heterogeneity of variance that existed from the early to the late trials and the skewness of the distributions on the late trials. The t tests that have been employed have either been corrected for heterogeneity of variance by the method suggested by Cochran and Cox (1) or the comparisons are between data that show reasonable evidence of meeting the underlying assumptions involved.

As a further precaution all statistical analyses involving percentages in this study were repeated using an arc-sine transformation of the percentages. In no case did the analyses using the transformed data result in significance levels appreciably different from those reported for the untransformed data.

accepted levels of statistical reliability and we cannot conclude that the two groups differ on the repetition of first-trial associations on subsequent trials.

A further confirmation of the lack of effect of insight upon the rate of avoidance learning is found when we compare the number of trials required by the insight and the noninsight groups to reach the criterion of two successive trials without a repetition of a punished response. The results of this analysis are seen in Table 2. As this table shows, the mean number of trials required for both groups is nearly identical and the t test falls far short of acceptable levels of significance.

Reaction times. While there is no evidence of differences in rate of avoidance learning between the insight and the noninsight *S*s, there are some marked differences between these two groups in reaction times to stimulus words that have had punished responses (critical stimuli). On each trial the median reaction time to critical and noncritical stimuli was computed for each *S*. In Fig. 2 the mean of these median reaction times as a function of number of trials is shown for the insight and the noninsight groups. For the noninsight group the mean reaction times for both the five critical and the ten noncritical stimuli decreases as a function of trials. By the last trial the reaction times are virtually the same for both critical and noncritical words.

For the insight group, however, the relation of reaction times to trials differs for the critical and the noncritical stimuli. The times for the noncritical words decrease with trials as did reaction times for both critical and noncritical words in the noninsight group, but for the critical stimuli there is an initial increase in reaction time. Only on the fourth trial does the mean reaction time to the critical words begin to decrease, and even then it is still longer than the reaction time on the first trial.

In order to evaluate the statistical reliability

TABLE 2

NUMBER OF TRIALS TO REACH THE CRITERION

Groups	Mean	Variance Estimate	t
Insight	4.18	2.8	.09
Noninsight	4.27	6.8	

Note.—The criterion trials have been omitted from the computations.

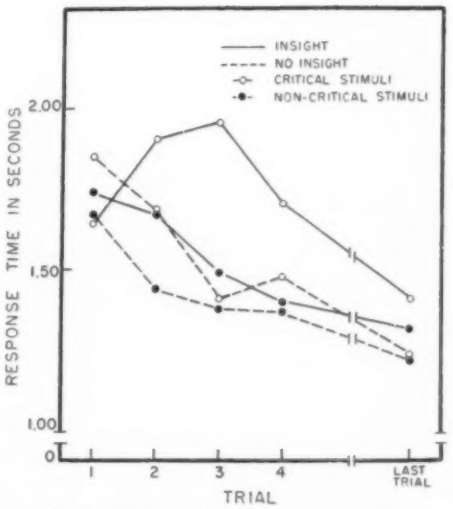

FIG. 2. REACTION TIMES TO CRITICAL AND NONCRITICAL STIMULI AS A FUNCTION OF TRIALS

of the interaction between insight, trials, and critical words that appears in Fig. 2, a modified four-way analysis of variance (insight, critical stimuli, trials, and *S*s) was performed. The analysis used was identical to that described by Lindquist as Plan VI (5). The data for this analysis were the median reaction times for critical and noncritical stimuli for each *S* on each of the first four trials.[5] The summary of this analysis is given in Table 3.[6]

Our main concern in this analysis was whether the trend of reaction times to the critical stimuli as a function of trials in the insight group was reliably different from the trend for the noncritical words and for both critical and noncritical words in the noninsight group. This question can be answered by the triple interaction between groups, critical stimuli, and trials. As is seen in Table 3, this interaction is highly significant ($p < .001$).

This difference in reaction time to critical and noncritical words between the insight and the noninsight groups provides us with an independent check upon the validity of classifying these *S*s as insightful or insightless on the

[5] Only the first four trials were used in this analysis since a number of the *S*s in both groups had reached the avoidance criterion by the end of the fourth trial.

[6] Bartlett's test for heterogeneity of variance did not reach significance at the .05 level.

basis of their verbal statements. The increase in reaction time to critical stimuli in the insight group is what might be expected in view of their statements that they deliberately withheld responses that had been punished while they attempted to think of new responses. It should also be noted that the reaction-time data bear out their statements that they soon learned to think of nonpunished responses "automatically" since the reaction times for the last trials closely approach those for the noncritical stimuli.

The analysis in Table 3 also confirms several other effects that are indicated in Fig. 2. Trials are, of course, a significant source of variation, but there is no over-all difference between the insight and the noninsight groups other than the interactions with other variables. The main effect attributed to critical and noncritical words is significant at the .05 level, suggesting that the reaction times to the critical words tended to be somewhat longer during the first four trials. The only other significant effect is the interaction between trials and the insight and noninsight groups. Inspection of Fig. 2 shows that for the insight group, reaction times decreased at a slower rate for the first three trials than did the reaction times for the noninsight group.

TABLE 3

SUMMARY OF THE ANALYSIS OF VARIANCE OF THE REACTION-TIME DATA FOR TRIALS 1–4

Source	Sum of Squares	*df*	*F* Ratio	*p*
Total	47.32	175		
Trials	1.46	3	.49/.07 = 7.00	<.001
Critical stimuli	1.49	1	1.49/.21 = 7.09	<.05
Insight	1.03	1	1.03/1.36 = .76	>.10
Subjects	27.13	20	1.36/.07 = 19.42	<.001
Trials × critical stimuli	.29	3	.10/.07 = 1.43	>.10
Trials × insight	.95	3	.32/.07 = 4.57	<.01
Critical stimuli × insight	.07	1	.07/.21 = .33	>.10
Trials × critical stimuli × insight	2.06	3	.69/.07 = 9.86	<.001
Within-subject error terms:				
Pooled subjects by trials	4.25	60	.07/	
Pooled subjects by critical stimuli	4.15	20	.21/	
Pooled subjects by trials by critical stimuli	4.44	60	.07/	

TABLE 4

COMPARISON OF GROUPS ON THE AVERAGE NUMBER OF CHAINED ASSOCIATIONS PER STIMULUS AND THE AVERAGE PERCENTAGE OF FIRST-TRIAL RESPONSES

Groups	Mean Number of Associations		Mean Percentage of First-Trial Responses Repeated	
	Stimuli		Stimuli	
	Critical	Noncritical	Critical	Noncritical
Insight	4.5	4.4	49	66
Noninsight	4.4	4.5	50	74

Chained associations. Following the avoidance conditioning on the word-association test, the *S*s were informed that there would be no more shocks in the experiment. They were then asked to chain-associate for 15 seconds to each of the stimulus words. In Table 4 we have presented a breakdown of the chained association data for the insight and noninsight *S*s in terms of mean number of associations and mean number of first-trial response words repeated as a function of critical and noncritical stimuli.

The data on total associations and on percentage of repetitions of the first-trial words were separately analyzed by a modified three-way classification analysis of variance (insight, critical words, and *S*s). The analysis for total associations revealed no significant effects. This is apparent from Table 4. However, the analysis of percentage of repetitions of first-trial responses resulted in one significant effect ($p < .01$). More first-trial responses to noncritical stimuli than to critical stimuli were elicited during chained associations. This finding indicates that the effects of shocks upon the *S*s' associations have persisted despite a somewhat different task and the conscious knowledge that electric shocks will no longer occur.

DISCUSSION

In order to explain the phenomena of repression in terms of the principles used in describing avoidance conditioning and the effects of punishment as Dollard and Miller (3) have done, it is necessary to show that implicit verbal behavior or thoughts are analyzable into stimulus-response sequences. It is also necessary to show that these stimulus-response sequences are learned and modified by the same principles of reinforcement that

govern other behavior. In their book, Dollard and Miller have given a plausible treatment of these points.

The results of the present experiment have contributed another step toward an analysis of repression in terms of avoidance conditioning. There are two characteristics of repression that must be demonstrated by any experimental analogue of the repression process. In order for repression to be effective, the anxiety-provoking thought must be prevented from occurring. Thus, the experimental analogue must provide reasonable evidence that the changes in behavior have occurred not only at the overt level but at the covert level as well. Also, since a prominent feature of repression, as clinically observed, is the *S*'s lack of awareness of the repression process, the experimental analogue must demonstrate that the anxiety-provoking thought is eliminated "automatically." That is, the *S* should not have an awareness of having deliberately and intentionally changed his thoughts.

The results of the present experiment seem to meet these two requirements quite well, especially in the case of *S*s in the noninsight group. Not only did these *S*s show a marked and rapid decrease in punished associations, but this dramatic change in their associations occurred without verbal awareness on their part. These *S*s were unable under intensive questioning to describe what they had done to avoid shocks, and they reported no awareness of having deliberately withheld associations in an attempt to think of a new association.

Fortunately we do not have to rely upon the verbal reports of these *S*s alone to support our belief that this avoidance learning occurred without awareness. A comparison of the reaction times for the insight and the noninsight groups lends confirmation to the *S*s' verbal reports. The insight *S*s who had clear verbal awareness of how to avoid shocks and who reported having tried deliberately to withhold punished associations and think up new ones, were found to show an initial increase in reaction times to the critical stimuli for the first three trials. This increase in reaction time is what would be expected to accompany the implicit behavior that these *S*s describe having experienced. For the noninsight group, on the other hand, reaction times to the critical stimuli show a monotonic decrease from the first trial to the last. The reaction times to the critical stimuli in this group are closely comparable to the reaction times to the noncritical stimuli throughout the experiment. The reaction-time data for the critical words in the noninsight group correspond quite well with what would be expected of an "automatic" repression process.[7]

While the verbalizations and the reaction-time data for the noninsight group strongly indicate that avoidance learning had occurred without awareness, these same data also suggest that the behavior changes occurred at the covert as well as at the overt level. Not only do the *S*s from the noninsight group report a lack of withholding of associations, but their reaction-time data also indicate that their overt and covert responses were identical. Had these *S*s implicitly responded with punished associations, it seems likely that their association times would have been appreciably longer to the critical stimuli than to the noncritical stimuli. Actually, however, there was little or no difference in the association times to the two classes of stimuli. Even among the *S*s in the insight group, the reaction times to critical stimuli had become comparable to the reaction times for noncritical stimuli by the last trial. This finding again corresponds with their verbal report that after a few trials the new associations "came to mind automatically" and they no longer thought of the punished responses.

One further piece of evidence that the shocks had effected associations on the implicit level is to be found in the data on the chained associations. Had these punished associations been present in implicit behavior, we would have expected more of them to have attained

[7] It is interesting to equate the performance of the insight group with the process of conscious suppression and the performance of the noninsight group with the process of repression as these processes are described by Dollard and Miller (3). In both groups the punished responses were successfully avoided, but in the noninsight group the avoidance occurred "automatically" without awareness and without an increase in reaction times. In the insight group, on the other hand, the avoidance was a deliberate conscious act and was accompanied by increased reaction times while these *S*s searched for substitute responses. In the noninsight group the data indicate that the thought itself or the implicit verbal response was initially conditioned while in the insight *S*s the overt verbalization was first conditioned and then the thought or implicit response.

overt expression when the threat of punishment was removed. Both groups, however, gave fewer of the punished responses during chained association than would be expected on the basis of the number of first trial responses obtained to the noncritical stimuli.

While it was not the purpose of the present experiment to contribute to the theoretical controversy concerning the manner in which punishment eliminates responses, our data seem to favor an interpretation along the lines of an inhibitory process or decrement in the absolute response strength of the punished responses. An increase in response strength of alternative responses alone does not seem adequate to account for the data obtained from chained associations. Here the *S*s had an opportunity to give a relatively large number of responses to each stimulus. Thus, it could be expected that not only would the stronger responses tend to occur, but there was the opportunity for responses lower in the response hierarchy to occur also. Yet only a few punished responses occurred, a result which is all the more surprising since these punished responses were for the most part initially the dominant response in the hierarchy.

SUMMARY

The *S*s were presented with a 15-item word-association list under the guise that they were taking part in an experiment to determine the limit of speed of associations. They were instructed to respond with the first word that came to mind as quickly as possible after the stimulus word was presented. During the first trial on the list, the *E* administered a strong electric shock immediately after five arbitrarily selected response words. The *S*s were then given a number of further trials on the same 15 stimulus words. Every time an *S* responded with one of the five first-trial punished responses, he received another electric shock. At the conclusion of this phase of the experiment, *S*s were informed that there would be no further shocks. They were then asked to chain-associate to each of the stimulus words.

On the basis of intensive questioning at the conclusion of the experiment, it was possible to classify the *S*s into insight and noninsight groups. The insight *S*s were characterized by having a high degree of verbal awareness as to the basis for the electric shocks and what they had to do to avoid receiving them. The noninsight *S*s had a low level of verbal awareness as to the reasons for the shocks and what they had done to avoid them. Both insight and noninsight groups showed a rapid and marked learning of avoidance behavior. No reliable differences were found between the two groups in terms of rate of learning or on the number of trials to achieve the criterion of two successive trials without the occurrence of a punished response. However, there were clear differences between the two groups in reaction times to stimulus words that had elicited shocked responses. The nature of these differences corresponded with what would be expected from the *S*s' verbalization. Both groups showed a significant decrement in the number of punished responses that occurred during chained associations. The relation of these data to a theory of repression was discussed.

REFERENCES

1. COCHRAN, W. G., & COX, GERTRUDE M. *Experimental designs*. New York: Wiley, 1950.
2. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.
3. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
5. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
6. SIDOWSKI, J. B. Influence of awareness of reinforcement upon verbal conditioning. *J. exp. Psychol.*, 1954, **48**, 355–360.

Received November 7, 1955.

THE EFFECTS OF VARYING COMBINATIONS OF AUTHORITARIAN AND EQUALITARIAN LEADERS AND FOLLOWERS[1]

WILLIAM HAYTHORN
The RAND Corporation

DON HAEFNER
PETER LANGHAM
University of Rochester

ARTHUR COUCH
Harvard University

LAUNOR CARTER
The RAND Corporation

RECENTLY attention has been focused on the relationships between the personality characteristics of group members and the nature of the leadership that emerges from the interaction of these individuals with each other and their tasks (3, 8). The behavior of leaders in groups is determined not only by their own personalities, but also by the underlying needs, role expectations, and values of the group members. Conversely, the behavior of group members is partly determined by the personality of the leaders, particularly in interaction with their own needs and expectations. Of particular interest are groups in which there is some modal similarity or dissimilarity of the leader's and followers' personalities, creating situations where the interaction of different needs, expectations, perceptions, and values can be studied systematically. The present study was designed to investigate a specific case of the interaction between leader's and followers' personalities in the determination of behavior in groups.

This paper is the second report describing the results of a study of the behavior of authoritarian and equalitarian personalities in small groups. The California F scale was used as the principal measure of the authoritarian personality (1). In the first paper (11), behavioral differences between groups composed of individuals homogeneously high on the F scale and individuals homogeneously low on the F scale were reported. An analysis was made of the differences between these groups with respect to rated behavioral traits, recorded behavioral acts, and responses to a Post-Meeting Reaction Sheet. Significant differences between high-F and low-F *S*s in the predicted directions were found, indicating that a more "democratic" group culture prevailed in the low-F groups, and the low-F groups were more effective in dealing with the task problem. Significant differences—more marked than those distinguishing the groups as a whole—were found between the emergent leaders of the two kinds of groups, indicating that a qualitatively different kind of leadership occurs in groups composed of personalities at opposite ends of the authoritarian-equalitarian dimension. Emergent leaders in the low-F groups were more sensitive to others, more effective leaders, more prone to making suggestions for action subject to group sanction, and less likely to give direct orders to others.

To permit experimental manipulation of the leadership structure, these same subjects were subsequently assembled with appointed leaders in the study reported here. Half of the groups were composed with appointed leaders who had F scores similar to those of the other group members, while in the other half the appointed leaders had scores at the opposite end of the distribution. This design permitted an investigation of the effects of assembling groups with leaders whose personalities were similar to those of other group members, as compared with groups whose leaders' personalities were unlike those of other members. The major hypotheses examined were as follows:

1. *Hypothesis 1.* F-plus *leaders* (*a*) engage in less equalitarian behavior than F-minus leaders; (*b*) are less concerned with group approval for their actions; (*c*) engage in more autocratic behavior; and (*d*) show less sensitivity to others than do F-minus leaders.

2. *Hypothesis 2.* F-plus *followers* (*a*) engage in less equalitarian behavior; (*b*) show less sensitivity to others; (*c*) are more satisfied with appointed leaders; (*d*) are less critical of their own group; and (*e*) are more submis-

[1] The work described in this paper was done under a contract between the U. S. Navy, Office of Naval Research, and the University of Rochester. Dr. Launor F. Carter was the responsible investigator for that contract.

sive toward appointed leaders than F-minus followers.

3. *Hypothesis 3. Under F-minus leadership,* followers (*a*) have more influence on group processes; (*b*) develop a less formal group structure; and (*c*) express greater differences of opinion than under F-plus leadership.

4. *Hypothesis 4. With F-minus followers,* leaders (*a*) are not likely to engage in as much autocratic behavior; and (*b*) are not as distinct from other group members as are leaders with F-plus followers. That is, leaders with F-plus followers are thought, when under pressure, to take decisive, directive action in order to get the job done, because of the expectations on the part of F-plus followers regarding the role of an appointed leader.

5. *Hypothesis 5. In groups composed of leaders and followers with similar F-Scale scores,* (*a*) followers are more secure; (*b*) followers are more motivated to achieve the common group goal; (*c*) followers are more satisfied with the appointed leadership; (*d*) there is less personality conflict within the group; (*e*) there is less conflict between the leader and group members; and (*f*) there is higher group morale and productivity than in groups composed of leaders and followers with dissimilar F-scale scores.

METHOD

Subjects. The subjects (*S*s) for the experiments consisted of (*a*) 32 male undergraduate students who were high on the F scale, "conservative" as judged by the Cattell Q_1 scale, and relatively normal as estimated by the MMPI; and (*b*) 32 male undergraduate students who had low F-scale scores, were liberal as determined by the Q_1 scale, and were also "normal" as judged by the MMPI. The mean F-scale scores for the two groups were 124.5 and 66.9, respectively, using the 30-item, 7-points-per-item scale described in the *Authoritarian Personality* (1).[2]

The task. The task given the groups was the discussion and composition of a short script involving a human relations problem. Complete instructions were read to the *S*s before they began working, to the effect that they were to help *E* in developing a test of human relations skills. They were to be shown a film of a human relations problem, after which they were to answer a number of questions about how the problem had been handled, how it should have been handled, etc. When they finished answering the questions as individuals, they were to discuss similar questions as a group and arrive at answers to which they all agreed. Following completion of this phase of the task, they were to compose dialogue for similar films. Once their script was completed, they were to record it on a magnetic tape recorder. After reading the instructions, *E* showed the film, then left the room. He did not return until the group had completed all phases of the procedure outlined in the instructions. A typed copy of the latter was left on the table with the *S*s.

Observation procedures. While *S*s worked on the task in the experimental room, observers watched through one-way vision mirrors from an adjacent room. Four *O*s (first- and second-year graduate students) recorded their observations in three ways, as follows:

1. While observing, *O*s categorized and "typed out" behavioral units as they occurred, using a Stenotype system of interaction recording described in an earlier paper (4). The category system was revised, however, to fit the needs of this study.

2. At the end of 30 minutes of recording, the pair of *O*s was relieved by another pair. Immediately after being relieved, *O*s rated each *S* on 16 behavioral characteristics, defined in the previous report (11), as follows: striving for individual prominence, friendliness, security, influence, equalitarianism, striving for group approval, submissiveness, striving for goal achievement, self-isolating behavior, effective intelligence, sensitivity to others, leadership, aggressiveness, autocratic behavior, social ability, and nonadaptability. These ratings had an average interobserver reliability of .75, with a range from .31 to .91. Since the group sessions varied in length from two to four hours, each *S* was rated by each *O* at least once in each group, and usually two or more times.

3. At the end of each group session, all four *O*s and *E* filled out a Post-Meeting Reaction Sheet (PMRS) designed to describe aspects of the group not easily attributed to single individuals. Examples of the items on the PMRS are:

a. This group was not serious enough—there was too much playing around.

b. The group effectiveness was interfered with by personality clashes between members.

There were 37 such items, to each of which *O*s indicated the extent of their agreement, using a 7-point scale.

4. In addition, *S*s were asked to fill out a very similar PMRS, and to rate the degree to which they were satisfied with their group, the extent to which they thought the group had been productive, to rank the group members in order of how well they were liked, and to rank group members in terms of their relative contributions to the achievement of group goals.

"Emergent" group sessions. The *S*s were scheduled to come to the laboratory work room in groups of four. In their first session—results of which are described in the previous paper (11)—each of the sixteen groups were composed of *S*s homogeneous with regard to F-scale scores. In this session, there was no experimental structuring of leadership.

Appointed leader sessions. After the emergent sessions, all individuals were reassembled with three members from their original emergent groups and another member from a different group. The alien member was appointed as leader (by telling him he was to take charge of group activities that day and would be responsible for getting the job done). In half of the groups

[2] Further details regarding the selection of subjects are given in Reference 11.

the leader had an F score similar to that of group members, and and in half it was dissimilar. Thus there four groups with F-plus leaders and F-plus followers, four with F-plus leaders and F-minus followers, four with F-minus leaders and F-plus followers, and four with F-minus leaders and F-minus followers.

In Session III, two of the original emergent-group members who had been together in Session II were reassembled with the member who had been withdrawn during Session II. In addition, an alien member was introduced as the leader for that session. Groups having had an F-plus leader in Session II were given an F-minus leader in Session III, and vice versa.

Characteristics of appointed leaders. The experimental design required that 32 of the 64 *Ss* be appointed as leaders, 16 in Session II and 16 in Session III. Two were chosen from each of the original groups as follows: The men in each group were assigned ranks on the basis of the leadership nominations made by *Ss* in their PMRS, and also on the basis of leadership ratings made by *Os*. Finally, in the F-minus groups each *S* was ranked on the basis of the equalitarian ratings given by *Os*, while in the F-plus groups they were ranked on the basis of ratings by *Os* of autocratic behavior. These three sets of ranks were averaged, and the two individuals (Ranks I and II) with the lowest average ranks were selected to serve as appointed leaders in subsequent groups. A later check showed that this procedure, designed to assure as nearly as possible that the appointed leaders would behave in an authoritarian or equalitarian manner, resulted in selecting the same leaders as would have been chosen by subject ratings alone if ties were broken by observer ratings.

Analysis of data obtained in the emergent sessions (11) indicated significant behavioral differences between the individuals chosen from the F-plus groups and those chosen from the F-minus. In summary, leaders chosen from the F-minus groups were rated as being more friendly, more secure, more equalitarian, less striving for group approval, more submissive, less self-isolating, showing more effective intelligence, showing greater sensitivity to others, and showing greater leadership. The pattern of results generally agrees with current opinion concerning the differences between authoritarian and equalitarian leaders, except that the differences in aggressiveness and autocratic behavior failed to reach statistical significance.

Observers were not informed of the composition of the groups during the experiment, and were unable to guess the composition pattern significantly more accurately than chance. This finding may be attributed to an observer expectation that authoritarian individuals would be more aggressive, argumentative, and generally uncompromising towards others. Instead, in the emergent sessions they were rated as being significantly more concerned with group approval than were the F-minus *Ss*. Such concern for group approval in groups having no formal authority structure is theoretically a central facet of the authoritarian personality structure (1), underlying the tendency to conform to social mores. However, our *Os* apparently tended to identify pleasant social behavior as evidence of equalitarianism and consequently made many errors in attempting to estimate group composition. The fact that analyses generally supported our hypotheses in spite of observer failure to distinguish the groups is presumptive evidence that results to be reported are not attributable to bias.

RESULTS

Trait ratings of appointed leaders. The experimental design, a 2 × 2 factorial design with leaders' and followers' personalities as the main effects, permitted testing of several sources of variance. Group means and results of the analyses of variance of ratings by *Os* are presented in Table 1. Only the ratings yielding significant *F* ratios are included, in order to conserve space. Seven of the 16 rating variables did not yield significant *F* ratios for any of the three sources on either session.

The data entered in Table 1 are mean ratings of the appointed leaders, averaged over four *Os*. The four groups are F-plus leaders with F-minus followers, etc. The within-group variance is taken as the best estimate of error, or variance not attributable to experimental conditions. The "effects on leaders" are divided into three sources: F scores of the followers, F scores of the leaders, and whether the leader and followers were homogeneous or heterogeneous with regard to F scores (the latter is the leader-by-follower interaction in the analysis of variance).

Trait ratings of followers. Analysis of the ratings of followers' behavior by *Os* parallels that for ratings of leaders. The main effects are the F scores of the followers and of the leaders. Again the interaction mean square indicates the effects of homogeneity vs. heterogeneity of leader-follower combinations. Only seven of the 16 variables yielded significant *F* ratios. These are presented in Table 2, with the group means and the within-group variance estimates.

There were three followers in each of the 16 groups in each session. Thus, each of the means of the four conditions are based on ratings of 12 individuals. However, in an earlier study (9) it was shown that the ratings of individuals within a group are not independent. Therefore, the average rating of the three followers in each group was computed and these averages used in computing means and within-group variance. Each condition mean is based on the averages of four groups. This procedure results in 15 degrees of freedom (16 − 1), distributed as follows: 1 for leaders' F scores, 1 for followers' F scores, 1 for interac-

TABLE 1

TRAIT-RATING RESULTS: APPOINTED LEADERSHIP SESSIONS

(Leaders)

Analysis of Variance		Leader Means					Effect on Leaders of:		
Trait names	Session	F+ leaders		F− leaders		s^2_w	F+ vs. F− followers *F* ratio	Leader's personality *F* ratio	Group's composition *F* ratio
		(F + F)†	(F− F)†	(F + F)	(F− F)				
I. Individual Prominence	II	4.76	3.19	4.08	4.13	0.16	14.574(F+)**	0.028(F−)	1.097(Homo)
	III	4.51	3.08	4.28	3.52	0.77	6.298(F+)*	0.056(F−)	0.572(Homo)
V. Equalitarianism	II	3.56	4.76	5.29	4.82	0.70	0.747(F−)	4.569(F−)	3.954(Hetero)
	III	3.27	4.04	4.89	5.48	1.00	1.837(F−)	9.350(F−)*	0.033(Hetero)
VI. Striving for Group Approval	II	3.89	3.62	4.49	4.26	0.25	1.039(F+)	6.264(F−)*	0.008(Homo)
	III	4.12	3.31	4.42	4.24	0.72	1.360(F+)	2.068(F−)	0.567(Homo)
VII. Submissiveness	II	3.03	3.28	3.43	2.75	0.59	0.325(F+)	0.026(F+)	1.462(Hetero)
	III	2.57	3.77	3.34	3.07	0.41	2.100(F−)	0.011(F−)	5.165(Hetero)*
IX. Aggressiveness	II	2.98	1.78	1.92	2.49	0.39	1.013(F+)	0.329(F+)	8.037(Homo)*
	III	2.95	1.63	1.96	1.34	0.51	7.441(F+)*	3.223(F+)	0.964(Homo)
X. Autocratic	II	3.64	1.99	2.21	2.61	0.56	2.773(F+)	1.150(F+)	7.459(Homo)*
	III	3.56	1.98	2.20	1.64	0.43	10.821(F+)*	6.877(F+)*	2.471(Homo)
XI. Effective Intelligence	II	4.32	4.46	5.23	4.92	0.31	0.095(F+)	5.963(F−)*	0.653(Hetero)
	III	4.21	3.90	4.96	5.42	0.53	0.038(F−)	9.759(F−)**	1.099(Homo)
XII. Sensitivity	II	3.27	4.26	4.89	4.47	0.77	0.423(F−)	4.361(F−)	2.607(Hetero)
	III	3.28	3.85	4.80	5.39	0.83	1.616(F−)	11.304(F−)**	0.001(Homo)
XV. Nonadaptability	II	2.01	1.38	1.31	1.70	0.14	0.407(F+)	1.055(F+)	7.469(Homo)*
	III	1.45	1.75	1.45	1.08	0.60	0.008(F+)	0.759(F+)	0.759(Hetero)

* p (.05) = 4.75: df 1/12.

** p (.01) = 9.33.

† (F + F), (F − F), refers to the personality of Followers, (F+ Followers) (F− Followers), respectively.

tion, and 12 for within-group variance. The estimate of significance so derived is believed to be very conservative.

Behavioral category indices for followers. It will be recalled that *O*s categorized and recorded behavioral acts as they occurred. Forty-three categories were used, but in order to simplify the analysis and increase the frequency of acts per category, these were combined into 17 category indices: I. Friendly acts; II. Positive affect acts; III. Direction-taking acts; IV. Self-emphasizing acts; V. Asking for suggestion or sanction; VI. Asking for group evaluation; VII. "Democratic" acts; VIII. Directive acts; IX. Expressions of confusion or lack of orientation; X. Hostile acts; XI. Negative affect acts; XII. Initiating and integrating acts; XIII. Diagnosing and clarifying acts; XIV. General task participation acts; XV. Self-isolating acts; XVI. Withdrawing and out-of-field activity; and XVII. Tension release and out-of-field activity. These indices are not independent, in that some categories of behavior were included in more than one index. For a more complete description of the categories and indices, the reader is invited to consult the earlier report (11). Results of the analysis of these indices are presented in Table 3.

Subject's Post-Meeting Reaction. At the end of each session, all four *S*s were asked to complete a Post-Meeting Reaction questionnaire, asking them to express the degree to which they agreed to a number of statements about the group. Agreement was expressed on a 7-point scale with the middle point missing to eliminate noncommital responses (except the items on motivation and productivity, which had 9 points). Items on the questionnaire were grouped on a priori grounds into indices, and the means of the three followers in each group were computed. These follower means were then analyzed in the same way as previous data. Results of this analysis for all

TABLE 2

TRAIT-RATING RESULTS: APPOINTED LEADERSHIP SESSIONS
(Followers)

Analysis of Variance		Group Means					Effect on Followers of:		
Trait names	Session	F+ followers		F− followers		s^2_w	F+ vs. F− leadership F ratio	Follower's personality F ratio	Group's composition F ratio
		(F+ L)†	(F− L)†	(F+ L)	(F− L)				
III. Security	II	3.90	4.54	4.10	4.25	0.17	3.678(F−)	0.046(F+)	1.428(Hetero)
	III	4.54	4.30	4.17	4.78	0.07	2.146(F−)	0.196(F−)	10.484(Homo)**
IV. Influence	II	3.85	3.91	3.92	3.81	0.09	0.028(F+)	0.014(F+)	0.310(Hetero)
	III	4.03	4.11	3.73	4.28	0.06	6.544(F−)*	0.262(F+)	3.671(Homo)
V. Equalitarianism	II	2.52	3.16	3.57	3.52	0.45	0.525(F−)	3.831(F−)	0.777(Hetero)
	III	2.64	2.61	2.94	3.99	0.32	3.211(F−)	8.711(F−)*	3.600(Homo)
VI. Striving for Group Approval	II	3.12	3.16	2.72	3.04	0.16	0.823(F−)	1.717(F+)	0.480(Homo)
	III	3.50	3.30	2.94	3.45	0.08	1.241(F−)	2.046(F+)	6.165(Homo)*
VIII. Striving for Goal Achievement	II	3.34	3.63	3.94	3.90	0.27	0.218(F−)	2.827(F−)	0.398(Hetero)
	III	3.90	3.84	3.90	4.64	0.13	3.628(F−)	4.881(F−)*	4.881(Homo)*
XI. Effective Intelligence	II	3.25	3.48	3.97	3.62	0.20	0.072(F+)	3.636(F−)	1.702(Hetero)
	III	3.42	3.44	3.47	4.25	0.11	5.589(F−)*	6.537(F−)*	5.244(Homo)*
XII. Sensitivity	II	2.83	3.07	3.67	3.36	0.42	0.001(F+)	3.093(F−)	0.726(Hetero)
	III	2.92	2.96	3.30	4.08	0.31	2.184(F−)	7.338(F−)*	1.780(Homo)

* p (.05) = 4.75: df ½.
** p (.01) = 9.33.
† (F + L), (F − L), refers to personality of leaders: (F + Leaders) (F − Leaders), respectively.

TABLE 3

BEHAVIOR CATEGORY INDICES: APPOINTED LEADERSHIP SESSIONS
(Followers)

Analysis of Variance		Group Means					Effect on Followers of:		
Index names	Session	F+ followers		F− followers		s^2_w	F+ vs F− leadership F ratio	F+ vs. F− personality F ratio	Group's composition F ratio
		(F+L)	(F−L)	(F+L)	(F−L)				
VI. Asking for Group Evaluation (14, 11)	II	1.28	1.86	2.83	2.58	0.66	0.16(F−)	7.87(F−)*	1.07(Hetero)
	III	2.87	1.73	1.93	2.60	3.18	0.07(F+)	0.00	1.03(Homo)
VII. "Democratic" Acts (34, 12, 14, 11)	II	2.24	2.67	3.71	3.68	1.11	0.14(F−)	5.54(F−)*	0.20(Hetero)
	III	5.00	2.78	3.22	3.37	6.85	0.62(F+)	0.20(F+)	0.82(Homo)
IX. Expressions of Confusion or Lack of Orientation (22, 33, 61)	II	1.43	2.28	1.58	1.40	0.56	0.81(F−)	0.96(F+)	1.91(Hetero)
	III	2.75	0.89	1.29	1.97	1.01	1.38(F+)	0.14(F+)	6.35(Homo)*
XIII. Diagnosing and Clarifying Acts (35, 38)	II	2.21	3.12	5.90	4.38	2.32	0.16(F+)	10.60(F−)**	2.56(Homo)*
	III	2.47	1.57	3.04	3.87	3.76	0.00	2.18(F−)	0.80(Homo)
XVI. Withdrawing and Out-of-field Activity (75, 76)	II	20.21	7.09	5.01	8.06	45.06	1.87(F+)	5.08(F+)*	5.18(Homo)*
	III	8.93	12.74	7.31	14.97	79.93	1.65(F−)	0.00	0.19(Homo)
XVII. Tension Release and Out-of-field Activity (8, 76)	II	51.74	21.24	11.45	11.92	258.73	3.49(F+)	9.51(F+) **	3.70(Homo)
	III	32.14	39.39	22.68	28.56	276.56	0.62(F−)	1.49(F+)	0.01(Hetero)

* p (.05) = 4.75: df ½.
** p (.01) = 9.33.

indices yielding significant F ratios are given in Table 4. No significant results were obtained for 12 of the 21 indices from the *S*s' PMRQ. These 12 were: II. Degree of equal participation; III. Degree of personality conflict; IV. Degree of informal friendliness; VII. Satisfaction with leadership; VIII. Degree of conflict within group; IX. Competence of members; X. Morale; XV. Formation of cliques; XVI. Personality conflict between leader and group members; XXI. Differences of opinion; XXII. Competition among members; and XXIII. Lack of cooperation.

Observers' Post-Meeting Reactions. Four *O*s also completed the PMRS. The same indices were computed, with the exception that *O*s answered questions regarding the formality of group structure and communication effectiveness, and did not answer the question concerning relative efficiency of leadership. Results of the *O*s' PMRS analysis are presented in Table 5. Nineteen of the 23 indices yielded at least one significant F ratio. Those that did not were: XVIII. Degree of goal-oriented control by appointed leader; XIX. Domination by appointed leader; and XXII. Competition among members.

Experimental findings. The results, in general, support the hypotheses stated at the beginning of the paper. With regard to Hypothesis 1, F-plus leaders were rated as being less equalitarian, less concerned with group approval, more autocratic, and less sensitive to others than F-minus leaders (Table 1). They were also rated as showing less effective intelligence (Table 1). The latter result is believed to be partially a function of the particular task given the groups, in this case a task requiring analysis of basic values which is presumably alien to F-plus personalities.

F-plus followers, as predicted from Hypothesis 2, were rated as being less equalitarian behaviorally, and as showing less sensitivity to others (Table 2). They were also rated as

TABLE 4

SUBJECTS' POST-MEETING REACTION SHEET INDICES: APPOINTED LEADERSHIP SESSIONS (Followers)

Analysis of Variance		Group Means					Effect on Followers of:		
Index names	Session	F+ followers		F− followers		s^2w	F+ vs. F− leadership F ratio	Follower's personality F ratio	Group's composition F ratio
		(F+ L)†	(F− L)†	(F+ L)	(F− L)				
I. Dissatisfaction with Goal Progress	II	3.13	2.30	2.52	2.92	0.23	0.81(F+)	0.00	6.53(Homo)*
	III	2.53	3.28	2.50	2.38	0.25	1.59(F−)	3.45(F+)	2.97(Hetero)
V. Definiteness of Leadership	II	5.17	5.35	3.98	4.00	0.10	0.42(F−)	62.18(F+)**	0.27(Hetero)
	III	4.40	4.31	4.33	3.64	0.10	6.03(F+)*	5.40(F+)*	3.70(Hetero)
VI. Striving for Equal Participation	II	5.96	5.75	4.42	5.21	0.36	0.95(F−)	12.06(F+)**	2.77(Homo)
	III	5.96	5.79	4.54	5.42	0.65	0.77(F−)	4.93(F+)*	1.67(Homo)
XI. Group Productivity	II	6.83	7.83	6.75	5.92	0.55	0.05(F−)	7.26(F+)*	6.10(Hetero)*
	III	7.25	6.67	6.92	6.75	1.15	0.49(F+)	0.05(F+)	0.15(Homo)
XIV. Motivation Towards Group Goal	II	5.50	5.17	3.83	4.00	1.62	0.02(F+)	4.96(F+)*	0.15(Homo)
	III	5.58	5.33	4.25	4.58	1.11	0.01(F−)	3.90(F+)	0.31(Homo)
XVII. Relative Efficiency of Leadership	II	5.50	5.50	6.08	4.25	0.68	4.94(F+)*	0.65(F+)	4.94(Hetero)*
	III	5.41	3.55	4.16	5.41	0.58	0.60(F+)	0.60(F−)	15.46(Homo)**
XVIII. Degree of Goal-oriented Control by Appointed Leader	II	4.83	6.04	5.44	4.94	0.39	1.28(F−)	0.64(F+)	7.43(Hetero)*
	III	5.56	4.41	4.69	5.00	0.48	1.34(F+)	0.16(F+)	4.10(Homo)
XIX. Domination by Appointed Leader	II	3.21	3.46	2.29	2.67	0.17	2.27(F−)	16.97(F+)**	0.09(Homo)
	III	2.79	2.72	2.42	2.04	0.68	0.27(F+)	1.50(F+)	0.12(Hetero)
XX. Satisfaction with Appointed Leader	II	4.58	5.67	4.50	3.50	0.78	0.01(F−)	6.45(F+)*	5.53(Hetero)*
	III	5.25	5.22	3.91	3.66	1.34	0.05(F+)	5.79(F+)*	0.03(Hetero)

* p (.05) F ($3/12$) = 4.75.

** p (.01) = 9.33.

† (F + L), (F − L), refers to personality of leaders (F + leaders), (F − leaders), respectively.

TABLE 5

OBSERVERS' POST-MEETING REACTION SHEET INDICES: APPOINTED LEADERSHIP SESSIONS (Groups)

Analysis of Variance		Group Means					Effect on "Groups" of:		
		F+ "Groups"		F− "Groups"			Followers		
Index names	Session	(F+ L)	(F− L)	(F+ L)	(F− L)	s^2w	F+ vs. F− leadership F ratio	F+ vs. F− personality F ratio	Group's composition F ratio
I. Dissatisfaction with Goal Progress	II	3.85	2.48	2.52	2.92	0.76	1.23(F+)	1.04(F+)	4.10(Homo)
	III	3.23	3.81	2.21	2.21	0.81	0.42(F−)	8.50(F+)*	0.42(Hetero)
II. Degree of Equal Participation	II	2.88	3.02	3.17	3.08	1.20	0.01(F−)	0.10(F−)	0.04(Hetero)
	III	3.38	3.48	2.28	3.78	0.33	7.71(F−)*	1.93(D+)	5.90(Homo)*
III. Degree of Personality Conflict	II	3.73	3.48	2.90	3.58	0.64	0.29(F−)	0.83(F+)	1.35(Homo)
	III	3.21	4.22	3.54	2.54	0.20	0.00	9.29(F+)*	20.59(Hetero)*
IV. Degree of Informal Friendliness	II	4.73	4.62	4.90	4.40	0.68	0.56(F+)	0.00	0.22(Hetero)
	III	5.00	4.47	4.32	5.48	0.36	1.11(F−)	0.31(F−)	7.98(Homo)*
V. Definiteness of Leadership	II	5.52	5.14	4.04	5.06	0.40	1.00(F−)	6.04(F+)*	4.93(Homo)*
	III	5.09	4.84	4.58	4.60	0.42	0.12(F+)	1.34(F+)	0.18(Homo)
VI. Striving for Equal Participation	II	3.40	4.85	2.88	3.53	0.75	7.81(F−)*	2.51(F−)	0.30(Hetero)
	III	3.98	4.72	4.32	5.78	0.89	5.43(F−)*	2.20(F−)	0.55(Homo)
VII. Satisfaction with Leadership	II	4.95	5.04	5.60	4.85	0.68	0.13(F+)	0.01(F−)	2.58(Homo)
	III	5.70	4.35	5.10	5.80	0.26	1.61(F+)	2.75(F−)	16.01(Homo)**
VIII. Degree of Conflict Within Group	II	3.26	3.29	2.69	4.09	0.73	2.77(F−)	0.07(F−)	2.58(Homo)
	III	2.81	3.76	3.18	2.38	0.40	0.06(F−)	2.64(F+)	7.71(Hetero)*
IX. Competence of Members	II	4.70	5.85	5.58	5.15	0.34	1.50(F−)	0.10(F−)	7.32(Hetero)*
	III	4.93	4.60	5.60	6.18	0.37	0.17(F−)	13.85(F−)*	2.30(Homo)
X. Morale	II	4.30	4.70	4.48	3.92	0.84	0.03(F+)	0.43(F+)	1.12(Hetero)
	III	4.80	4.00	4.45	5.72	0.21	1.06(F−)	9.04(F−)*	20.68(Homo)**
XI. Group Productivity	II	6.15	7.35	6.25	5.50	1.60	0.16(F−)	1.81(F+)	2.26(Hetero)
	III	6.20	5.15	6.50	7.90	0.78	0.16(F−)	11.99(F−)**	7.74(Homo)*
XII. Formality of Group Structure	II	3.10	2.25	2.50	2.40	0.44	0.97(F+)	1.29(F+)	0.46(Homo)
	III	2.85	2.30	2.90	2.00	0.34	6.26(F+)*	0.19(F+)	7.69(Homo)*
XIII. Communication Effectiveness	II	5.40	5.40	5.50	4.90	0.49	0.74(F+)	0.33(F+)	0.74(Hetero)
	III	5.40	4.95	4.75	5.95	0.35	1.59(F−)	0.35(F−)	7.69(Homo)*
XIV. Motivation Towards Group Goal	II	3.65	5.15	4.90	5.05	0.90	3.03(F−)	1.47(F−)	2.03(Hetero)
	III	3.80	4.00	4.55	5.95	0.44	5.75(F−)*	16.38(F−)**	3.24(Homo)
XV. Formation of Cliques	II	2.90	3.60	2.90	3.45	1.32	1.18(F−)	0.02(F+)	0.02(Hetero)
	III	1.90	3.20	2.30	2.10	0.30	3.99(F−)	1.62(F+)	7.42(Hetero)*
XVI. Personality Conflict Between Leader and Group Members	II	2.72	3.15	2.33	3.03	0.84	1.52(F−)	0.30(F+)	0.08(Homo)
	III	2.12	3.83	3.05	2.15	0.23	2.86(F−)	2.41(F+)	29.38(Hetero)**
XX. Satisfaction with Appointed Leader	II	3.15	2.25	3.55	4.35	1.06	4.37(F−)	0.06(F−)	0.29(Homo)
	III	3.35	4.00	3.50	5.65	1.52	5.15(F−)*	2.13(F−)	1.48(Homo)
XXI. Difference of Opinion	II	3.15	3.70	2.60	4.65	0.87	7.76(F−)*	0.18(F−)	2.58(Homo)
	III	2.75	3.62	3.22	2.52	0.43	0.07(F−)	0.92(F+)	5.82(Hetero)*
XXIII. Lack of Cooperation	II	4.05	3.25	3.10	3.75	1.19	0.02(F+)	0.17(F+)	1.77(Homo)
	III	3.25	4.30	3.70	2.30	0.41	0.30(F+)	5.82(F+)*	14.55(Hetero)**

* p (.05) F (3/2) = 4.75.
** p (.01) = 9.33.

showing less goal motivation (Table 2). Again, the latter results were probably highly contingent on the task. The category indices supported these results by indicating that F-plus followers engaged in less asking for group evaluation and "democratic" acts than did F-minus followers (Table 3).

F-plus followers were also found to be more satisfied with appointed leaders than F-minus followers, and were apparently less critical of their own groups in the sense that they rated their groups as more productive and more motivated to achieve the group goal than did the F-minus followers (Table 4). That this result is due to differential response tendencies is suggested by the fact that *O*s rated the groups with F-minus followers as showing more group productivity and goal striving (Table 5), and recorded more withdrawing and out-of-field activity for F-plus followers (Table 3).

Hypothesis 2*e*, that F-plus followers would be more submissive toward appointed leaders, was not supported by *O*s' ratings of submissiveness nor by the recorded "direction-taking" behaviors. There was some support for this hypothesis from the PMRS analyses, however, where F-plus followers rated their groups as having more definite leadership and as being more dominated by the appointed leader (Table 4), and *O*s rated groups with F-plus followers as having more definite leadership (Table 5).

It was hypothesized (3*a*) that under F-minus leaders, followers—whether they were F-plus or F-minus—would exert more influence on group processes. This prediction was supported by ratings of "influence" and "effective intelligence" (Table 2). The hypothesis is further supported by the finding that *O*s rated groups with F-minus leaders as having a significantly higher degree of equal participation (Table 5).

Hypothesis 3*b*—that groups with F-minus leaders would develop a less formal group structure—was supported by the results with *O*s' PMRS rating (Table 5). By "formal group structure" *O*s attempted to describe the extent to which the leader played his task role in a stilted manner. This finding does not imply any differences in the level or definiteness of leadership. Rather, it seems that the crucial differences here are in the style in which authoritarian and equalitarian personalities display leadership.

Hypothesis 3*c*, that followers would express greater differences of opinion under F-minus than under F-plus leaders, was supported by *O*s' ratings on the PMRS (Table 5), but not by the *S*s' PMRS ratings and not by the behavioral category index of "negative affect acts," which subsumes the category "disagrees." The category index results were in the predicted direction, but did not approach statistical significance. Hypothesis 3*c* is, therefore, considered of doubtful tenability.

Hypothesis 4 stated that leaders with F-minus followers would differ from those with F-plus followers by (*a*) engaging in less autocratic behavior, and (*b*) being less distinct from other group members. These predictions were both supported by findings that leaders with F-plus followers were rated higher on autocratic behavior (Table 1), that *S*s and *O*s both rated groups with F-plus followers as having more definite leadership (Tables 4 and 5), and that F-plus followers rated their groups as being more dominated by the appointed leader (Table 4). (The latter results were offered as support for Hypothesis 2*e* above. Their repetition here is indicative of the interactive nature of the hypothesis investigated.)

Hypothesis 5*a*—that followers are more secure in homogeneous groups—was supported by the difference in rating of "security" by *O*s (Table 2). Followers in homogeneously composed groups were also rated as "striving for goal achievement" more than those in heterogeneous groups (Table 2), in support of Hypothesis 5*b*.

The hypothesis that followers would be more satisfied with the appointed leaders in homogeneous groups was not supported. In fact, quite the opposite was found, i.e., followers in the heterogeneous groups rated themselves as significantly more satisfied with their appointed leader. No satisfactory explanation of this result has been achieved, but inspection of the means indicates that the largest differences in satisfaction were between F-plus followers with F-minus leaders and F-minus followers with F-minus leaders. The latter were relatively quite dissatisfied with

their leaders. To what extent this finding is due to differential response tendencies cannot be determined with the data obtained.

Hypothesis 5*d*—that there is less personality conflict in homogeneous groups—was supported by *O*s' PMRS ratings (Table 5), but not by *S*s' ratings. *O*s' ratings are believed to provide better tests of hypotheses involving *overt* behavior (because of the broader framework in which the *O*s were indoctrinated before the experiment), and the hypothesis is tenable within that context. However, if "personality conflict" is taken to imply subjective reactions, the hypothesis is not supported. High-F persons are probably less sensitive to personality clashes than are low-F, and probably are not as likely to perceive conflicts between themselves and others as are low-F persons.

The *O*s' ratings of conflict between leaders and group members support the hypothesis that there is less such conflict in homogeneous groups (Table 5). Again, this difference was not found with *S*s' ratings of their own groups.

Finally, ratings by *S*s and *O*s yielded conflicting and inconclusive results with regard to hypothesis 5*f*, that homogeneous groups are more productive and have higher morale. With regard to productivity and motivation, nothing very tenable can be concluded. From the observers' point of view, however, it appears that the homogeneous groups had higher morale, higher communication effectiveness, less conflict, and better cooperation than the heterogenous groups (Table 5).

DISCUSSION

The results, then, support most of the hypotheses tested, indicating significant differences for each of the five major comparisons. Two of these—the comparison of F-plus and F-minus leaders and the comparison of F-plus and F-minus followers—are rather conventional types of comparisons. They simply indicate that the behavior of individuals can be predicted, to some extent, from measures of those individuals' attitudes or personality characteristics.

The other three major comparisons, however, are quite different. Each of these suggests that to some extent the behavior of individuals in groups is dependent on the personalities of other group members. We found, for example, that there were differences between leaders with F-plus followers and leaders with F-minus followers. This finding indicates that the behavior of leaders is, to a significant degree, a function of the attitudes or personality characteristics of the followers. Conversely, the behavior of followers is found to be significantly a function of the attitudes or personality characteristics of the leaders. The latter, of course, is almost a *sine qua non* for the concept of leadership, but the former has received less attention theoretically and almost none empirically.

Finally, the significant interactions indicate that differences between F-plus and F-minus leaders are contingent on whether the followers are high or low F, and vice versa. This result supports the so-called interaction theory of leadership (8) even more forcefully than the observations referred to in the preceding paragraph. Separating the effects of followers on leader behavior from the interaction effects directs attention to the mutually dependent, adaptive nature of small-group behavior, and appears to facilitate our ability to predict.

The cautious reader will have noted the authors' tendency to reject the null hypothesis occasionally on tenuous bases. There are two relatively noticeable kinds of inconsistencies in the results—those between Session II and Session III, and those between *O*s' and *S*s' ratings. These have been resolved in the discussion by accepting significant results in one session if the other session was not highly contradictory, and to accept *O*s' ratings as more descriptive of the overt behavior and *S*s' as more descriptive of the subjective feelings involved.

Previous evidence (5) has indicated that appointing an individual to a position of leadership in laboratory groups results in his behaving differently from other group members. Another study (10) has shown that group characteristics can be predicted from behavioral and personality measures of individual group members. The present study indicates that conclusions drawn from these previous studies must be tempered by considerations of the interaction among group members, and particularly that between leaders and other group members. All of these findings, of course, are probably contingent on such specifics as the kind of task con-

sidered, the population studied, and the nature of the leadership appointment.

SUMMARY

Four-man groups were required to perform on a task in which *S*s discussed specific human relations problems presented by film, then composed and recorded dialogue for similar problems. One of the four men in each group was appointed as leader. The California F scale was administered to all *S*s, and group composition was varied systematically to yield four conditions: (*a*) high-F leaders with high-F followers; (*b*) high-F leaders with low-F followers; (*c*) low-F leaders with high-F followers; and (*d*) low-F leaders with low-F followers. Observations of behavior were recorded by: (*a*) ratings of individuals on 16 behavioral characteristics, (*b*) an interaction recording technique using a classification of behavior acts into 43 categories, and (*c*) responses by *O*s and *S*s to a Post-Meeting Reaction Sheet involving questions about the group.

Five major comparisons were made: (*a*) F-plus leaders with F-minus leaders; (*b*) F-plus followers with F-minus followers; (*c*) leaders of F-plus followers with leaders of F-minus followers; (*d*) followers with F-plus or F-minus leaders; and (*e*) groups in which leader and followers were homogeneous as compared to groups in which they were heterogeneous with regard to F-scale scores. For each comparison, significant differences were found consistent with a number of hypotheses drawn from the theory of the "authoritarian personality."

Results indicate that F-plus appointed leaders differ behaviorally from F-minus leaders, and that F-plus followers differ from F-minus followers. These two comparisons are conventional tests of behavioral differences related to personality characteristics. In addition, interaction analyses indicated that the behavior of leaders was a function of whether their followers were F-plus or F-minus, and the behavior of followers depended on the F-scale scores of their leaders. Several significant individual and group differences were found between heterogeneously and homogeneously composed groups. The implications of these results for an interaction theory of leadership, and for the prediction of behavior from personality measures of individuals, are discussed.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J. & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. BLUM, R. *The study of groups.* Washington, D. C.: Human Resources Research Office, Staff Memorandum, 1 August 1953.
3. CARTER. L. F. Leadership and small-group behavior. In M. Sherif & M. W. Wilson (Eds.), *Group relations at the crossroads.* New York: Harper, 1953.
4. CARTER, L. F., HAYTHORN, W., MEIROWITZ, BEATRICE, & LANZETTA, J. Note on a new technique of interaction recording. *J. abnorm. soc. Psychol.*, 1951, **46**, 258–260.
5. CARTER, L. F., HAYTHORN, W., SHRIVER, BEATRICE, & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589–595.
6. CATTELL, R. B., SAUNDERS, D. R., & STICE, G. *Handbook for the sixteen personality factor questionnaire.* Champaign, Ill.: Institute for Personality and Ability Testing, 1949.
7. DORRIS, R. J., LEVINSON, D. J., & HANFMANN, EUGENIA. Authoritarian personality studied by a new variation of the sentence completion technique. *J. abnorm. soc. Psychol.*, 1954, **49**, 99–108.
8. GIBB, C. A. The research background of an interactional theory of leadership. *Aust. J. Psychol.*, 1950, **2**, 19–42.
9. HAYTHORN, W. The influence of the individual group members on the behavior of co-workers and on the characteristics of groups. Unpublished doctor's dissertation, Univer. of Rochester, 1952.
10. HAYTHORN, W. The influence of individual members on the characteristics of small groups. *J. abnorm. soc. Psychol.*, 1953, **48**, 276–284.
11. HAYTHORN, W., COUCH, A., HAEFNER, D., LANGHAM, P., & CARTER, L. F. The behavior of authoritarian and equalitarian personalities in small groups. *Hum. Relat.*, 1956, **9**, 57–74.
12. STOGDILL, R. M. Personal factors associated with leadership: survey of the literature. *J. Psychol.*, 1948, **25**, 35–71.

Received November 15, 1955.

PARADOXICAL EFFECTS IN THE REACTION TIME OF SCHIZOPHRENICS[1]

P. H. VENABLES AND J. TIZARD

Medical Research Council, Unit for Research in Occupational Adaptation, Institute of Psychiatry, London, England

IT HAS long been established that for normal, practiced subjects, reaction time to a visual stimulus decreases as the intensity of the stimulus increases (1, 2, 5). Similar findings have been reported for auditory stimuli (3, 9). Hull (6, p. 68), summarizing a number of studies of different kinds, claims that there is a "definite tendency, other things approximately equalised, for trained responses to have a greater reaction potential as the stimulus increases in intensity, whether tested by reaction latency, amplitude, or intensity. Accordingly the principle may be considered as qualitatively well established."

Although among psychologists there has been much discussion about the specific quantitative laws which govern the functional relationship between intensity of stimulus and speed of response, the existence of such a relationship is not disputed. Wundt (12), it is true, is reported by Woodworth (11, p. 118) as saying that RT increases again at high intensities "because one can never be fully ready for a very intense stimulus"; but Woodworth adds that this statement has never been adequately tested. For the intensities usually employed in RT studies, no such increase has been reported.

Pavlov (8) showed, however, that dogs with "weak, inhibitory nervous systems," in whom pathological disturbances of the cortex are induced by means of functional interference, tend to respond to conditioned stimuli in a manner qualitatively different from that of healthy animals. Among the pathological phenomena described by Pavlov are "paradoxical" effects (in which weak stimuli call forth a larger conditioned response than do strong stimuli), and "ultraparadoxical" effects (in which excitatory stimuli are converted into inhibitory ones and inhibitory stimuli into excitatory ones). Stimuli of sufficient strength to bring abou paradoxical or ultraparadoxical effects are said to be of "ultramarginal" strength. Pavlov claimed that schizophrenic withdrawal and negativism are examples of paradoxical and ultraparadoxical phenomena.

Such evidence as there is to support Pavlov's claims is either clinical or based on analogy with experimental neurosis in animals. No experimental studies with human *S*s appear to have been carried out. If, however, paradoxical phenomena occur commonly in schizophrenia, it would seem to follow that they should be experimentally demonstrable in a reaction-time study using stimuli of different intensities. Strong stimuli, which ordinarily result in brisk responses by healthy people, should be "ultramarginal" for schizophrenic, and RTs should be slower with such stimuli than with weaker ones.

In a pilot study (10) in which four stimulus lights of different intensities were employed, 21 out of 24 schizophrenic *S*s showed such "paradoxical" increases in RT when reacting to lights of higher intensity in contrast to a group of normal *S*s whose RTs decreased when they responded to light stimuli of higher intensity. In the present study the factors affecting this paradoxical phenomenon are further investigated under more strictly controlled conditions, with additional *S*s and a larger number of stimulus intensities to cover the higher values more fully than had been done before.

METHOD

Procedure. The *S*s were tested individually under standard conditions. Eight intensities of stimulus light were used in the experiment, 20 consecutive responses being made to each light.

The stimulus consisted of a ground glass screen 0.75 inches in diameter, illuminated from the rear by a 150 watt bulb fed from a stabilized voltage source. The light source was set in a 2 x 2 feet matt black panel placed at an angle to the vertical so as to be at right angles to the *S*'s line of sight when seated at a distance of 2 feet from the panel. Variations in stimulus intensity were made by inserting a series of neutral density filters behind the ground glass screen. The intensities of the stimuli were measured by means of a photometer; the values obtained are given in Table 1. The background illumination reaching *S*'s eye was 1.7 foot candles.

[1] This work was carried out in Springfield Hospital, London, by kind permission of the Medical Superintendent, Dr. H. C. Beccle.

TABLE 1

VISUAL STIMULUS INTENSITIES IN FOOT CANDLES AND LOG FOOT CANDLES

Intensity	A	B	C	D	E	F	G	H
Foot candles	16	60	135	275	450	600	800	1500
Log foot candles	1.20	1.78	2.13	2.44	2.65	2.78	2.90	3.18

The *S*s were tested in two sessions, lights of four of the eight intensities being presented in the first session, and the remaining four in a second session which followed within 24 hours. Order of presentation was varied in a balanced latin-square design such that *S*s responded to lights of intensity A, C, E, and G in one session, and to B, D, F, and H in the other. Although 20 RTs were measured for each stimulus intensity, only responses 6–15 were used in the analysis to avoid warm-up and end-spurt effects.

A forewarning buzzer signal preceded by 3.0 seconds the onset of the stimulus light. The *S* reacted to the stimulus light by lifting his finger. The light then went out and was prevented by a relay mechanism from immediately reappearing when *S*'s finger was replaced on the key to await the next presentation of the stimulus light. If *S* failed to respond, the light remained on for 3.0 seconds. The cycle of buzzer and light was repeated once every 8 seconds. Between each group of 20 responses, a 3-minute rest period was allowed. The timing of the presentation was controlled electronically to an accuracy of 5 per cent. Measurement of each response was made by means of a cold cathode counter tube timer to the nearest millisecond.

Subjects. All patient *S*s were male nonparanoid schizophrenics between the ages of 21 and 52 years. One group consisted of 24 men (mean age, 38.8, *SD*, 8.7 years) who had been in hospital for longer than two years and who could be considered as displaying marked schizophrenic signs. Six out of the 24 had acted as *S*s in the pilot study carried out several months previously (10). Eleven other of these *long-stay* patients selected to participate did not cooperate. A second group consisted of 16 convalescent patients (mean age, 28.5, *SD*, 3.6 years) who had been in hospital less than one year, and recently received treatment. This *short-stay* sample was in general much less disturbed than the first group. No *S* had undergone a leucotomy or was receiving physical treatment or drugs at the time of the experiment.

As controls, four male nurses were also tested. As the pattern of their responses was similar to that already described as typical of normal *S*s it was not considered necessary to test a larger control group.

RESULTS

A preliminary analysis showed that the patterns of results obtained from schizophrenic *S*s on the first occasion of testing were dissimilar to those obtained on the second. The results from each occasion were therefore analyzed separately. Table 2 shows the mean RTs for the 12 long-stay patients (Group A), who on the first occasion of testing responded to lights of intensity A, C, E and G, and similar data for the 12 long-stay patients (Group B), who responded to lights of intensity B, D, F and H.

It is seen from these data that, in general, the pattern in both cases is similar, being one of a decrease in RT from the weakest to the next intensity of stimulus (C or D), followed by a subsequent increase in RT with each further increase in stimulus intensity. The decrease in RT from intensity E to intensity G (Group A) is due to the excessively long reaction to intensity E given by two of the 12 *S*s in this group. When the data from these two *S*s are omitted, the means for this group show a lengthening in RT with each increase in stimulus intensity after the second.

The rise in RT with increase in intensity from C to G is shown by 11 out of 12 *S*s, and the rise in RT with increase in intensity from D to H is also shown by 11 out of 12 *S*s. Using the sign test (4) as a measure of significance in view of the skewed distribution of the data, the rise in RT with increase in stimulus intensity in each of the groups of 12 *S*s is significant at the .01 level. The decrease in RT with increase in intensity from A to C and from B to D is not significant.

Since a similar pattern was found in both groups of 12 *S*s, it was considered legitimate to combine the data from each group for the first occasion of testing in order to show how responses varied over the whole range of eight stimulus intensities. As the mean RTs for Group A were higher than those for Group B, those of Group A were proportionately reduced, so that the mean RTs of each group

TABLE 2

MEAN RTS OF LONG-STAY AND SHORT-STAY SCHIZOPHRENIC AND NORMAL *S*s ON THE FIRST OCCASION OF TESTING

Subjects	Stimulus Intensity							
	A	B	C	D	E	F	G	H
Long-stay schizophrenics								
Group A	.72		.57		.94		.86	
Group B		.66		.57		.77		.98
Combined group	.69	.66	.55	.57	.91	.77	.83	.98
Short-stay schizophrenics								
Combined group	.34	.35	.31	.32	.30	.34	.32	.31
Normals								
Combined group	.24	.22	.24	.21	.22	.19	.21	.21

were similar. This was done by multiplying the mean RTs of Group A by .965, this figure being derived by division of the summated mean score of Group B by that of Group A. The mean RTs thus obtained for the whole range of intensities presented on the first occasion of testing are shown in Table 2.

The data show that there is a fall in RT from the lowest intensity to the intensities C and D, after which RT rises to a higher figure with increase in intensity.

With the group of short-stay, less severely ill patients, the results obtained for reaction to stimuli of all intensities on the first occasion of testing are shown in Table 2. In view of the closeness of the mean scores of the two subgroups, the one reacting to lights of intensity A, C, E, and G, and the other to B, D, F, and H, no proportional adjustment was necessary. For this group, there were no significant trends in RT with different stimulus intensities.

The mean response times for the four normal *Ss* for the first occasion of testing are also given in Table 2. Response times for the second occasion of testing were closely similar. They are much shorter than those obtained from the schizophrenics, and like those of the short-stay schizophrenics, they do not show a paradoxical rise with increasing stimulus intensity. The normal group shows only a slight fall in RT as stimulus intensity increases. A more marked fall was not seen because at the intensities employed in the present study, the slope of the curve of performance was approaching the asymptote.

On the second occasion of testing, when the remaining four intensities in the second half of the latin-square design were used, there was no statistically acceptable evidence with either the long- or the short-stay schizophrenic *Ss* of a change in RT with the increase in stimulus intensity for C and D and beyond. Mean response times over all intensities on the second occasion do not differ significantly from those on the first, although the previously found pattern of response times is absent.

The mean RTs for the long-stay group and for the short-stay group of schizophrenics on the second occasion are given in Table 3.

The results for the six long-stay schizophrenics who had taken part in the pilot study some months before did not differ noticeably from those of the other 18 long-stay patients.

TABLE 3

MEAN RTs OF LONG-STAY AND SHORT-STAY SCHIZOPHRENIC *Ss* ON THE SECOND OCCASION OF TESTING

Subjects	Stimulus Intensity							
	A	B	C	D	E	F	G	H
Long-stay schizophrenics	.76	.74	.65	.84	.81	.75	.77	.76
Short-stay schizophrenics	.39	.37	.37	.31	.35	.42	.35	.32

In the present experiment five out of the six *Ss* showed a paradoxical effect on the first but not on the second occasion of testing.

There was no evidence of order effects either within or between testing sessions in either group of patients. Independence of the results in relation to age is shown by a non-significant correlation between age and increase in RT between intensities C and H.

To examine in more detail the nature of the paradoxical phenomenon found with long-stay schizophrenics on the first occasion of testing, further analysis of the data was undertaken. King (7) pointed out that slow RTs are characteristically found among more seriously ill schizophrenics. Though no ratings of degree of illness were made in the present study, the data were examined to see if the paradoxical effect would be shown more markedly by those schizophrenics with the slowest RTs. The total group of long-stay schizophrenics was divided on the basis of RT to the weakest intensity into slow, medium and fast subgroups containing equal numbers of *Ss*. Linear regression constants, of the form $y = ax + b$, to cover the range from intensity C to intensity H were calculated for each of these subgroups, as well as for the short-stay schizophrenics and the normal controls. These values are given in Table 4.

Table 4 shows that for the slow long-stay schizophrenic *Ss*, the slope of the gradient of response is steep and positive ($a = +0.872$), but it decreases regularly through a low positive value for the short-stay schizophrenics to a negative value for the normals as the level of RT at intensity C decreases. The negative regression slope shown by the normal *Ss* in this experiment is in accord with the findings of other workers (1,2,5). The close relationship of regression slope to level of performance at the weakest stimulus intensity is also shown by a correlation of .74 ($p < .001$)

TABLE 4

REGRESSION EQUATIONS SHOWING TRENDS OF PERFORMANCE OF SCHIZOPHRENIC AND NORMAL *Ss*

Long-stay schizophrenics	
Slow	RT = +.872 log i + .85
Medium	RT = +.223 log i + .46
Fast	RT = +.141 log i + .32
Short-stay schizophrenics	RT = +.003 log i + .31
Normals (male nurses)	RT = −.029 log i + .23

between these two variables based on all the 44 *Ss* in the study.

A qualtitative comment can be made. Some patients were tested who were too ill to cooperate in one or both testing sessions. Their results therefore could not be included in those given above. It was observed however that some of these *Ss* found it particularly difficult to respond to the brighter lights. Up to the cessation of the light stimulus after the elapse of the 3-second "on" period, they appeared to be pressing down on the key; they were only able to release it when the light went off. Unfortunately, no pressure recordings were made during the experiment, so there is no quantitative confirmation of this impression.

DISCUSSION

The results show with a reasonable degree of clarity that with chronic schizophrenic *Ss*, the predicted paradoxical effect of an increase in RT with an increase in stimulus intensity beyond an optimum point occurs on initial testing. A problem arising from these data, however, is that the paradoxical effect does not appear on the second occasion of testing. A possible explanation is that on the first occasion, in an unfamiliar situation, *Ss* were apprehensive and in a state of considerable emotional excitation. The excitation resulting from the stimulus lights, superimposed on this general excitation, became "ultramarginal" in strength and resulted in the paradoxical effect observed. On the second occasion of testing, *Ss* were more familiar with the situation, and the total excitation did not reach ultramarginal strength. This dependence of the paradoxical effect on the *S*'s level of excitation is in line with Pavlov's views on the phenomenon (8, pp. 54–56), and it is predictable from his theory that after extensive testing when the general level of excitation of the *S* is very low because of boredom or fatigue, the paradoxical effect reappears. If subsequent experimentation should prove this to be the case, the second occasion of testing could probably be regarded as an intermediate state between two "phasic" states.

That the paradoxical effect was also shown in five out of six patients who had previously taken part in a pilot study might be considered to weaken the explanation suggested above. But even for these *Ss*, the situation was probably still a relatively unfamiliar one. The apparatus used differed somewhat from that which had been used in the pilot study; *E* was hardly known to the *Ss*; and a considerable time had elapsed since their previous testing.

That the phenomenon is fairly well estabilshed is shown by its appearance in a total of 56 out of 64 chronic schizophrenics tested in the pilot study (10) already mentioned, and the present investigation.

SUMMARY

A reaction-time experiment was carried out, using lights of different intensities, to test Pavlov's suggestion that paradoxical effects are commonly found in schizophrenics. It was predicted that whereas normal persons tend to respond more quickly to more intense stimuli, schizophrenics respond more slowly to such "ultramarginal" stimuli than to lights of weaker intensity. In 22 out of 24 chronic, nonparanoid schizophrenics, RTs to the brighter stimuli were longer than those to weaker stimuli in a first testing session but not in a second. Convalescent, short-stay schizophrenics and normal controls showed no paradoxical effects. A tentative hypothesis to explain why the chronic patients did not show paradoxical effects on retest is proposed.

REFERENCES

1. BERGER, G. O. Über den Einfluss der Reizstarke auf die Dauer einfacher psychisher Vorgange mit besonderer Rucksicht auf Lichtreize. *Phil. Stud.*, 1886, **3**, 38–93.
2. CATTELL, J. M. The influence of the intensity of the stimulus on the length of the reaction time. *Brain*, 1886, **8**, 512–515.
3. CHOCHOLLE, R. Variation des temps de reaction auditifs en fonction de l'intensité a diverses frequences. *Année psychol.*, 1945, **41–42**, 65–124.

4. DIXON, W. J., & MOOD, A. M. The statistical sign test. *J. Amer. statist. Ass.* 1946, **41,** 557–566.
5. FROEBERG, S. The relation between magnitude of the stimulus and the time of the reaction. *Arch. Psychol.*, 1907, No. 8.
6. HULL, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, **56,** 67–76.
7. KING, H. E. *Psychomotor aspects of mental disease.* Cambridge: Harvard Univer. Press, 1954.
8. PAVLOV, I. P. *Conditioned reflexes and psychiatry* W. H. Gantt (Trans.). London: Lawrence & Wishart, 1941.
9. PIÉRON, H. De la loi de variation des temps de latence en fonctions des intensités excitatrices par les sensations auditives. *C.R.Soc. Biol.*, 1919, **82,** 1116–1118.
10. VENABLES, P. H., & TIZARD, J. The effect of stimulus light intensity on the reaction time of schizophrenics *Brit. J. Psychol.*, 1956, **47,** 144–145.
11. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.
12. WUNDT, W. *Grundzuge der physiologischen Psychologie.* (5th Ed.) Leipzig: Englemann, 1911.

Received November 15, 1955.

REACTION TIME AND RESPONSE AMPLITUDE AS A FUNCTION OF ANXIETY AND STIMULUS INTENSITY

ALFRED CASTANEDA

Iowa Child Welfare Research Station, State University of Iowa

ACCORDING to Hull's (2) formulation concerning the joint influence of different drives, total drive strength is assumed to be composed of both irrelevant and relevant drives. Two rather specific consequences are theoretically expected as the result of variations in the magnitude of either type of drive. An increase in one, other factors constant, should result in greater total drive strength. However, increments in total drive strength attributable to an increase in the strength of one are not simply additive but depend on the existing strength of the other. For example, while total drive strength increases with increased irrelevant drive, the increment in total drive strength due to an increase in relevant drive is expected to be *relatively* greater, the lower the existing strength of the irrelevant drive. The net effect of increasing relevant drive strength under two levels of irrelevant drive, then, is to reduce the difference in total drive strength between the two levels of irrelevant drive. If the situation is relatively simple, then differences in the magnitude of total drive strength should be reflected in the amplitude and latency of the response.

The results of an experiment relevant to these theoretical implications were recently reported by Wenar (8). On the assumptions that differences in intensity of the stimulus in a simple reaction-time task should produce differences in relevant drive, and that anxiety, as inferred from scores on the Taylor Scale of Manifest Anxiety (7), contributes to irrelevant drive strength, the following hypotheses were formulated and tested: (*a*) the speed of reaction of anxious *S*s should be higher than that of nonanxious *S*s, and (*b*) the increase in speed of reaction attributable to increased stimulus intensity should be relatively greater for the nonanxious *S*s, thus reducing the difference between the anxious and nonanxious *S*s. Results confirming only the first of these two expectations were obtained, i.e., the speed of reaction of the anxious *S*s was found to be higher in comparison to those of the nonanxious *S*s, but the difference between them tended to remain relatively constant with increased stimulus intensity.

The present study presents further data relevant to these theoretical expectations. Both the amplitude and the speed of reaction in a simple reaction-time task are used as bases for comparing anxious and nonanxious *S*s.

METHOD

Subjects. The *S*s were 44 high school boys and girls selected on the basis of scores made on the Taylor scale from a total sample of 152. The distribution of scores for these 152 *S*s has been reported by Lyle and Smock (3), who found it to be comparable to that reported by Taylor with over 1,900 college students. The anxious group consisted of 22 *S*s, 11 boys and 11 girls, whose scores fell within the range 19–34. The nonanxious group was composed of 22 *S*s, 11 boys and 11 girls, whose scores fell within the range 2–12.

Apparatus. The apparatus consisted of a rectangularly shaped plywood box approximately 18 × 12 × 6 in. The response element, the handle and stirrup of a Smedley-type hand dynamometer, projected from the lower right-hand corner of the box and was the only part of the dynamometer visible to *S*. The CS, a buzzer, was provided by means of a direct current relay energized with alternating current. Variations in the loudness of the buzzer were made possible by means of a variable transformer. Fifteen volts were used for the weak intensity and 45 volts, which produced a fairly raucous sound, for the strong intensity. When the circuit was energized, the buzzer and a Standard Electric clock were simultaneously activated. The indicator on the face of the dynamometer, controlled by the spring-action of the dynamometer, exerted sufficient pressure at the zero position on the leaf of a leaf-type microswitch to keep the circuit closed. Exertion of from four to five kilograms of pressure by *S* resulted in sufficient deflection of an indicator to open the microswitch and hence break the circuit to the buzzer and the timer. Another indicator remained at the point of maximum pressure exerted by *S* and had to be returned manually to the zero position by *E*. Measures of the amplitude of *S*'s reaction as well as of its speed were obtained by this arrangement.

Procedure. The *S*s were told that the purpose of the experiment was to measure the speed of their reactions and that they were to squeeze the dynamometer as soon as they heard the buzzer. No information was given regarding the fact that the amplitude of their reactions was to be measured. An interval of seven seconds was used between *E*'s saying "Ready" and the presentation of the buzzer. The *E* reported *S*'s reac-

tion time to the nearest 1/100 second after every presentation. The *S* was required to keep his hand on the dynamometer at all times.

The stimulus was presented a total of ten times, five at the weak intensity and five at the loud intensity. The two intensities were randomely distributed over the ten presentations with the one restriction that no one intensity appear more than twice in succession.

RESULTS

Table 1 presents the results based on the amplitude scores by intensity, anxiety level, and sex separately. Table 2 presents a summary of an analysis of variance based on the data presented in Table 1. With the exception of the interaction between sex and anxiety, all other effects are significant.[1] Inspection of Table 1 reveals that the amplitude scores of the anxious *S*s were higher than those of the nonanxious, a finding consistent with the expectation of greater response strength due to greater irrelevant drive strength. The significant triple interaction, however, when viewed in conjunction with the data presented in Table 1, indicates that the expected greater increase in response strength for the nonanxious group with increased intensity held only for the boys, whereas the performance of the girls was contrary to expectation, i.e., the anxious girls showed the greater increase. Note should be taken of the slight but nonsignificant decrease in amplitude for the anxious boys with increased intensity.

Although the distribution of the latency measures did not appear to be unduly skewed, a reciprocal transformation was performed on all reaction-time scores in order to make them comparable to Wenar's. Hence, the response measure may be referred to as speed of reaction rather than reaction time.

Table 3 presents the data based on the speed-of-reaction scores by intensity, anxiety levels, and sex. On the basis of these data, it is apparent that the speed-of-reaction scores tended to be higher for the nonanxious group at the weak intensity. The increase in speed of reaction associated with increased intensity appeared to be greater in the case of the anxious group. The relatively greater increase in speed of reaction for the anxious group, despite the tendency for their scores to be lower in comparison to the nonanxious group at the weak intensity, was great enough to result in higher speed-of-reaction scores for these anxious *S*s at the loud intensity.

Table 4 presents a summary of the analysis of variance performed on the data presented in Table 3. The results of this analysis indicates that with the exception of the main effects of anxiety and the interaction between sex and anxiety, all other effects were statistically significant. Tests of the simple effects of

[1] While differences meet the appropriate tests of statistical significance, it should be noted that there is considerable overlap in the distributions and that the results are not of a magnitude to yield a high degree of accuracy of prediction with respect to individual *S*s.

TABLE 1

RESPONSE AMPLITUDES IN KILOGRAMS TO THE WEAK AND LOUD BUZZER FOR THE ANXIOUS AND NONANXIOUS BOYS AND GIRLS

	Boys				Girls			
	Weak		Loud		Weak		Loud	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Anxious	28.18	5.56	27.27	4.77	17.68	4.52	20.18	3.14
Nonanxious	23.41	8.32	26.50	6.61	15.68	6.02	17.18	5.43

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE BASED ON AMPLITUDE SCORES

Source	*df*	MS	*F*	*p*
Anxiety	1	153.02	10.25	<.005
Sex	1	1649.66	110.49	<.001
Sex × Anxiety	1	.31	—	
Error (b)	40	14.93		
Intensity	1	52.65	28.77	<.001
Intensity × Anxiety	1	12.27	6.71	<.01
Intensity × Sex	1	21.22	11.59	<.001
Intensity × Anxiety × Sex	1	17.70	9.67	<.005
Error (w)	172	1.83		

TABLE 3

MEAN OF SPEED-OF-REACTION SCORES AT THE WEAK AND LOUD BUZZER FOR ANXIOUS AND NONANXIOUS BOYS AND GIRLS

	Boys				Girls			
	Weak		Loud		Weak		Loud	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Anxious	5.98	1.20	10.29	8.42	4.94	1.60	5.35	1.29
Nonanxious	6.68	2.38	7.54	3.18	4.97	.69	5.08	.62

TABLE 4
SUMMARY OF ANALYSIS OF VARIANCE BASED ON SPEED-OF-REACTION SCORES

Source	*df*	MS	*F*	*p*
Anxiety	1	7.25	2.36	>.10
Sex	1	141.48	46.08	<.001
Anxiety × Sex	1	4.53	1.48	>.20
Error (b)	40	3.07		
Intensity	1	44.62	15.38	<.001
Intensity × Anxiety	1	19.29	6.65	<.025
Intensity × Sex	1	29.87	10.30	<.005
Intensity × Anxiety × sex	1	13.69	4.72	<.05
Error (w)	172	2.90		

anxiety, however, did not prove to be significant.

DISCUSSION

Although the results based on the amplitude measure alone tend to be consistent with the assumption of anxiety as an irrelevant drive, they do not lend themselves to an unequivocal interpretation concerning the manner in which stimulus intensity and anxiety combine to determine total drive strength. While the amplitude scores of the anxious group did tend to be higher than those of the nonanxious group at both intensities, the expected reduction in this difference at the loud intensity occurred only in the case of the boys, whereas increased intensity produced an increase in this difference among the girls. On the other hand, the expectation of greater response strength attributable to anxiety, at both intensities, was not reflected in the speed measure and, contrary to expectations, the increase in response strength attributable to increased intensity was relatively greater for the anxious group. If one assumes that the reaction-time task involved only the single motor reaction, then the results based on the speed measure alone are not consistent with the major premises of the present study. It is possible, however, that the task did involve more than the single motor reaction. One may assume that the overt motor reaction was not directly elicited by the CS but was mediated by some implicit verbal response made directly to the CS. In the present experiment, for example, since *S* was instructed to "squeeze" at the onset of the buzzer, the word "squeeze" may have served to mediate the overt motor reaction of squeezing. If such was the case, it is possible that any other implicit verbal tendencies, e.g., those aroused by the drive stimuli associated with anxiety, could have competed with this mediating response, thereby delaying its occurrence. Assuming the occurrence of the motor reaction to be dependent on the prior occurrence of the mediating response, the effects of other competing tendencies would be manifested in a slowing of the motor reaction. On the other hand, since total effective drive was still operative at this time, the amplitude of the mediated reaction should reflect the magnitude of drive. The present results did show anxiety to be consistently associated with greater amplitude.

In this connection, it should be mentioned that Spence (5) and Postman and Kaplan (4) have stressed the point that measures of speed can provide a sensitive index of the presence of competing tendencies even when other measures fail to do so. If, in addition, it is assumed that the effect of increased stimulus intensity was to increase the strength of the mediating response *relative* to other competing tendencies which may have been present, it would be expected that the weaker the intensity, the greater the tendency for the presence of other competing tendencies to be reflected in the speed measure. By the same token, however, the tendency for differences in the magnitude of drive to be reflected in the speed measure should be greater with increased intensity due to the relative increase in the strength of the mediating response over other competing tendencies. The results reported here with respect to speed are in accord with these speculations. A somewhat similar proposal has been advanced by Spence, Farber, and Taylor (5), who suggest that the difference between anxious and nonanxious *S*s in classical conditioning situations may be a function of the degree of noxiousness (intensity) in the experimental situation.

Although it might be argued that the present results are consistent with the speculations of an implicit mediating verbal habit and the presence of competing habits associated with the drive stimuli of anxiety, there are no attributes of the present experiment nor of the results which offer any substantially *direct*

evidence in their support. Also, they appear to imply that habits associated with the drive stimuli of anxiety are in some sense incompatible with the mediating *verbal* habit. Such reasoning suggests that some of these interfering habits are also verbal in character. This suggestion accords with the observation that anxious persons are frequently distractible and unable to concentrate, particularly on tasks which require manipulation or retention of symbolic (verbal) material. In fact, the digit-span task is frequently considered as sensitive to the influence of anxiety. Such notions imply rather strong or perseverative implicit verbal habits associated with anxiety, and it is interesting to note that the Taylor scale contains a number of items concerning inability to concentrate, preoccupation with failure, worry, etc. Manifest anxiety symptoms of this type may involve such persistent implicit verbal habits. On the other hand, there are a number of items concerning sweating, heart palpitations, headaches, tension, etc., many of which could be interpreted to indicate that some of the habits associated with anxiety are chiefly somatic or nonverbal. If anxious *S*s could be differentiated reliably as to the type of habit associated with anxiety, i.e., whether they are mainly of the verbal or somatic type, their performance might be expected to vary as a function of the type of habit. For example, should the task require a high degree of symbolic manipulation, anxious *S*s essentially of the somatic type might perform better than anxious *S*s of the predominantly verbal sort. Farber (1) has discussed in detail this stimulus function of anxiety. The present speculations are in line with his considerations in that they suggest a method for the analysis of the habits associated with the stimulus component of anxiety.

SUMMARY

The present study was concerned with reaction time and response amplitude as a function of CS intensity and anxiety (Taylor scale). Five presentations of two auditory stimuli varying only in intensity were given to each of 22 anxious and 22 nonanxious *S*s. Increased intensity resulted in significant increases in amplitude and speed of reaction. Mean amplitude scores of the anxious group were significantly greater than those of the nonanxious group. A significant interaction was found between anxiety and intensity, based on a tendency for the speed of reaction of the anxious group, in comparison to the nonanxious group, to be slower at the weak intensity but faster at the strong intensity. Significant sex differences were found on both measures. The results were discussed in terms of competing implicit response tendencies.

REFERENCES

1. FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.
2. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
3. LYLE, W. H., & SMOCK, C. D. The relationship between manifest and test anxiety. *Proc. Iowa Acad. Sci.*, 1955, **62**, 405–410.
4. POSTMAN, L., & KAPLAN, H. L. Reaction time as a measure of retroactive inhibition. *J. exp. Psychol.*, 1947, **37**, 136–145.
5. SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *Kentucky symposium, Learning theory, personality theory, and clinical research*. New York: Wiley, 1954. Pp. 1–21.
6. SPENCE, K. W., FARBER, I. E., & TAYLOR, ELAINE. The relation of electric shock and anxiety level to performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.
7. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
8. WENAR, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, **49**, 335–340.

Received November 16, 1955.

A GENETIC APPROACH TO THE INTERPRETATION AND EVALUATION OF THE PROCESS-REACTIVE DISTINCTION IN SCHIZOPHRENIA[1]

WESLEY C. BECKER

University of Illinois

THE behavior disorders currently labeled "schizophrenia" pose some of the most complex problems psychologists and psychiatrists have yet to unravel. The problem facing the investigator in this area remains an overwhelming one. If controlled research is to be undertaken, some meaningful ways of organizing and understanding the gross individual differences among schizophrenics is necessary.

One compelling suggestion comes from the many prognostic investigations by such people as Hunt and Appel (13), Langfeldt (16), Kant (14), Becker and McFarland (5), Kantor, Wallner, and Winder (15), Stotsky (25), Benjamin (9), and Wittman (26, 27, 28). All of these prognostic studies have led to the conclusion that remitting schizophrenics have a better personality adjustment before and during illness than nonremitting schizophrenics. In addition, these and other studies[2] have suggested two consistent clusters of signs having differential prognostic significance. Briefly, the signs related to poor prognosis, which for convenience have been termed the "process syndrome," are (*a*) a "shut-in," withdrawn, inadequate prepsychotic personality, (*b*) slow, insidious development of psychosis, (*c*) relative absence of precipitating factors, and (*d*) presence of dull, rigid, or inappropriate affect. The signs related to good prognosis and termed the "reactive syndrome" are (*a*) relatively normal prepsychotic personality, (*b*) acute onset of psychosis, (*c*) presence of identifiable precipitating factors, and (*d*) presence of strong emotionality or tension. One implication of these findings is that a variable of severity of illness or level of adjustment is important to understanding individual differences among schizophrenics.

The consistency of the prognostic findings has led some to postulate the possibility of two kinds of schizophrenia with possible different etiologies. The process syndrome is most frequently assumed to have an "organic" basis, and the reactive syndrome is assumed to have a "psychological" basis (7). However, there are many reasons for rejecting such a conclusion. First, examination of the research data reveals considerable group overlap. Consequently, any attempt to force all schizophrenics into one group or the other would be clinically difficult and arbitrary. Also, if one recognizes with Bellak that schizophrenia is a deficit reaction which may be brought about by any combination of 40 or more etiological factors (8), and if one accepts the fact that twenty years of research have failed to find clear etiological differences between any subgroupings, then the conception of two types of schizophrenia loses usefulness.

An alternative assumption is that the process-reaction syndromes are best thought of as end points on a continuum of levels of personality organization. "Levels of personality organization" is a difficult concept to define succinctly. It is concerned with changes in the content and structure of mental organization as the human organism develops toward psychological maturity. A complete definition would encompass such factors as objectivity in perception (constancy), differentiation of needs, interests, and other aspects of personal motivation, and the degree of emotional control or adaptive functioning under stress. Lewin (17), Baldwin (2), and especially Werner (29) have all attempted to deal with this construct. Common to most attempts to conceptualize levels of personality is the general idea that "the development of biological forms is expressed in an *increasing differentiation* of parts and increasing *subordination*, or *hierarchization*" of the parts with respect to the whole (29). While space

[1] Condensed from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Stanford University. The writer is indebted to Dr. C. L. Winder, Dr. Quinn McNemar, and Dr. Albert Bandura for their guidance throughout this research. He also wishes to thank Dr. Leonard Ullman for his assistance in scoring part of the data, and the director and staff of Agnews State Hospital, Agnew, California, for their helpful cooperation in the execution of this research.

[2] The reader is referred to Bellak (7) and Kantor *et al.* (15) for reviews of these studies.

does not permit a detailed analysis of the parallel, the general proposition that the process syndrome reflects a very primitive, undifferentiated personality structure and the reactive syndrome a more highly organized one offers a potentially fruitful point of departure.[3]

To demonstrate the organizing value of this interpretation of the process-reaction concept, and to facilitate measurement of the dimension of personality levels, the following hypothesis was generated and tested: Schizophrenic patients who more nearly approximate the process syndrome show more regressive and immature thinking processes than schizophrenics who more nearly approximate the reactive syndrome. The argument here involves taking a prognostic rating as the criterion measure, placing each schizophrenic *S* along a process-reaction dimension. If this measure is essentially a measure of level of personality organization, and if the hypothesis is sound, then maturity of thought processes should be predictable from the ratings.

METHOD

Newly admitted schizophrenics were interviewed and their clinical records examined in order to gather social history data. This information was then evaluated in terms of the process-reaction dimension. Thinking processes were evaluated from psychological test results. Finally, interrelations between history and test data were explored.[4]

Subjects

The *S*s were 51 diagnosed schizophrenics selected at time of admission to a California state hospital (Agnews) on the basis of their being testable, non-Negro, under 41 years of age, without complicating neurological or other physical disorder, and on the basis of their being either a first admission case or a second admission case where the first admission was under four months duration. Twenty-four *S*s were male, 27 female. Average age for the men was 29.4 and for the women, 31.1. Twenty per cent had attended college, and over 60 per cent had been graduated from high school. About half were married. All but seven cases were first admissions. Twenty-four were diagnosed as paranoid type, nine as acute undifferentiated type, eight as chronic undifferentiated type, eight as schizo-affective type, and one each as catatonic and hebephrenic.

Evaluation of Thinking Processes

The independent variable to be evaluated is degree of regressive and immature thinking. In the literature one finds many definitions of schizophrenic thinking disorder, but few, if any, of such definitions contain the possibility of scaled quantification. The conceptual structure taken in this paper, however, suggests both theoretical and empirical (i.e., studies with children) methods of scaling severity of thinking disorder by initially equating it with immaturity of thinking.[5] To

[3] A more detailed theoretical exposition can be found in Becker (6). Reference to Werner (29) will also help to make this parallel more explicit for the reader not familiar with Werner's work.

[4] In considering the method and results which follow, the reader should be aware of a definite methodological difficulty. The author made both the case history and test evaluations. The time-consuming nature of the study did not make it possible to obtain an independent rater for one of the sets of variables. However, definite steps were taken to guard against contamination. Previous studies with the Rorschach have indicated that the scoring system used could be very reliably rated. In addition, an independent reliability study of the Proverbs scoring system was planned within this study. Thus, it seemed unlikely that bias would enter the evaluation of test data, or if it did with regard to the Proverbs, such bias would be detected through a higher correlation with history data by the author than by the independent rater. No such effect was found. It seemed reasonable to assume that if any contamination occurred, it would be more likely in the more complex history ratings. To minimize any such possible effect, those few subscales of the Elgin Prognostic Scale which required direct rating of patient behavior were rated before testing. In addition, all 51 cases were interviewed and tested before other case-history evaluations were made. In the meantime, no attempt whatsoever was made to evaluate the test data. For this reason, by the time the author was ready to make the final evaluations of case histories from the clinical records—a time lapse of from one to four months from testing—it was close to impossible to associate a specific test protocol with a specific case history. While one cannot assert that no bias entered into the ratings, the rather high reliabilities of all measures used and the precautions taken make it unlikely.

Ideally a follow-up study using independent evaluations would be desirable. Such a study is being planned under the direction of C. L. Winder as part of a comprehensive five-year study. Because of the length of time before the results of Dr. Winder's study will be available and because the author believes this paper has definite methodological and theoretical contributions to offer which are independent of the results, publication was sought at what may seem in some ways to be a premature date.

[5] To avoid misunderstanding, elaboration of this point is needed. What is asserted here is that the *formal organization* of mental processes of severe schizophrenics and young children, or of less severe schizophrenics and slightly older children is quite similar. The assumed basis for this similarity is that the schizophrenic has been fixated at an immature level of personality organization and/or has reverted to simpler levels of organization under stress. This approach provides a rational framework for the initial scaling of

demonstrate the method and to test the hypothesis given above, the Rorschach and Benjamin Proverbs were used to evaluate thinking processes. The 1937 Stanford-Binet Vocabulary test was included in order to estimate verbal intelligence.

Rorschach. Administration of the Rorschach followed the procedure proposed by Beck (3, 4). The method of analyzing the Rorschach was derived most directly from Werner's developmental theories (29) and the empirical studies by Friedman (10), Siegel (23), and Hemmendinger (12). In studying the problem of perceptual regression in schizophrenia, Friedman developed a scoring system for the Rorschach which would presumably reflect levels of perceptual development as suggested by Werner's theory. In comparing adult normals, schizophrenics, and children of various ages on his measures, he found strong support for the hypothesis that the structural aspects of schizophrenic perception were in many ways similar to, though not identical with those of children. Werner's conceptual structure and these supporting studies encouraged the notion that a scale could be devised which would differentiate levels of mental organization within the schizophrenic population.

In quantifying his data, Friedman used two gross categories—genetic-early and genetic-late perceptions. Each genetic group was further divided into three progressive levels, giving a total of six levels. The empirical findings of Friedman and Siegel, as well as Werner's theories were employed in making these differentiations. In order to derive an over-all score to reflect average level of mental organization, each response was given a weight from one to six, corresponding to the genetic level of response, and the sum of the weights was divided by the number of responses.

A summary outline of the Rorschach scoring system with examples is given in Table 1.[6] Rationale for placement of categories is discussed below.

Level one is characterized by diffuse, global, undifferentiated perceptions. *Wa* is a direct indicator of the diffuse, global nature of perception at this level. *W*— indicates the lack of differentiation of inner and outer worlds and the syncretism in mental organization. *DW* is a clear example of a type of diffuse perception Werner has described by the phrase "*pars pro toto*"—any part has the quality of the whole (29). The contaminated and fabulized responses reflect both the concrete and syncretic nature of primitive perception. In these responses there is an absurd fusion of percepts on the basis of spatial identity and spatial contiguity. Perseveration is assumed to reflect the dynamic rigidity of the personality at this level. Empirically, Friedman, Siegel, and Hemmendinger found these perceptual classifications most characteristic of children under five years of age and of severely regressed schizophrenics.

Level two is characterized by attempted differentiation in which the diffuse and syncretic nature of perception is still apparent. Friedman found the *Da*, *DdD*, and *Dv* responses all relatively rare in children, suggesting that where perceptual organization is so advanced as to permit discreteness (*D*), the diffuse and syncretic modes of perception have already begun to wane. In psychopathology, however, one might expect to find a greater frequency of this combination of discreteness with diffusion and syncretism where there is regression from higher levels of perceptual differentiation. A similar explanation is offered by Friedman to account for the differences in relative frequencies of the *Wv* and *Dv* responses. The greater frequency of vague responses in the schizophrenic group may be the result of regressive diffuseness and vagueness interfering with percepts which might have been *Wm* responses. Because *Wv* shows some integrative effort with consideration of the formal aspects of the blot it is placed at level three. *Dv* lacks integrative effort and is therefore placed at level two. *D*— and *Dd*— again are indicative of an unsuccessful attempt at differentiated perception.

Level three is characterized by the achievement of fair differentiation with only rudimentary integrative efforts. Developmentally, the scores at this level (with the exception of *Wv*) are most characteristic of children from ages seven to ten. In some ways *Adx-Hdx* responses are like the *Dm* of the next level. However, they are placed at this lower level because they clearly indicate a failure at integration where inregration is usually easily achieved. *Dd* responses were omitted from Friedman's schema. However, Hemmendinger (12) and Siegel (23) indicate that *Dd*'s are found most frequently between the ages of six and ten. Within this age range, an average of 13 per cent of the responses were of the *Dd* variety. None was found before the age of six, and normal adults gave but 4 per cent *Dd*. Siegel's less severe paranoid group gave 15 per cent *Dd*, whereas his more severe hebephrenics and catatonics gave but 3 per cent. These data are consistent with the interpretation that *Dd* is most characteristic of a level of development where the primary focus is analytical. With Beck's tables as a guide (3, 4), *Dd*+ was placed at level three, while the more immature *Dd*— was placed at level two.

Level four is a stage of accurate differentiation with the ability to make simple integrations. The Mediocre responses scored at this level indicate the ability to meet certain constant typical requirements in form necessary for adult perception.

Level five is indicative of clearly integrative activity with the ability to subordinate differentiated parts to the whole.

schizophrenic mental productions. Empirical findings, such as those by Friedman (10), can then be used to refine such a scale and to take into account any *differences* which may exist in the mental productions of children and schizophrenics of a given organizational level.

[6] The reader is referred to Friedman (10) and Siegel (23) for more complete definitions of scoring categories, with the exception of Unusual Details and Perseveration. The Unusual Detail scores are discussed in the text. Friedman's use of a perseveration score based on Beck's content classes did not prove to be discriminatory between age groups. For this reason the writer adopted Piotrowski's (21) more severe criterion for perseveration. Examples in the table are from Friedman (10), Phillips and Smith (20), and the author's own experience.

TABLE 1

DEFINITIONS AND EXAMPLES FOR THE RORSCHACH GENETIC-LEVEL SCORING SYSTEM

Level	Classification	Definition	Examples
1	Amorphous Whole (*Wa*)	Shape plays no determinable role.	I. "Black paint" II. "Fire and smoke"
	Minus Whole (*W*−)	Content requires specific form not provided by blot.*	I. "A fly" IV. "Starfish"
	Confabulatory Response (*DW*)	A single detail is basis for interpretation of the whole.	VI. "Cat," because of "whiskers"
	Contaminated Response (*Con R*)	Fusing of two interpretations of the same blot area.	VI. "Turtle-skin rug"
	Fabulized Combination (*Fab C*)	Absurd combination on basis of spatial contiguity.	X. "Rabbit with worms coming out of eyes"
	Perseveration (*Per*)	Same content to 3 or more cards with little regard to form requirements.	I, IV, V. "Spider" VIII, IX, X. "Internal organs"
2	Amorphous Detail (*Da*)	Analogous to *Wa*.	II. (*D* 2) "Fire" VIII. (*D* 6) "Flesh"
	Confabulatory Detail (*DdD*)	Analogous to *DW*.	VI. (*D* 3) "Cat's head," solely on "whiskers"
	Minus Detail (*D*−)	Analogous to *W*−.*	II. (*D* 2) "Kittens"
	Vague Detail (*Dv*)	Form element is so unspecific that almost any blot area could encompass content.	II. (*D* 2, 3) "Blood stains" X. (*D* 9) "Island"
	Minus Unusual Detail (*Dd*−)	Analogous to *W*−.*	VI. (*Dd* 25) "Pig's foot"
3	Vague Whole (*Wv*)	Analogous to *Dv*.	I. "Piece of a puzzle" X. "Design," "Map"
	Oligophrenic Detail (*Adx-Hdx*)	Response to part of an *A* or *H* percept usually seen as a completed figure.	III. (*D* 6) "Head of a person" V. (*D* 4) "Wing"
	Plus Unusual Detail (*Dd*+)	Content is a reasonable match to blot area isolated.*	X. (*D* 26) "Funny face"
4	Mediocre Detail (*Dm*)	Form implied in outline and articulation matches blot area. At level of "populars."	III. (*D* 3) "Bow tie" X. (*D* 15) "Little bird"
	Mediocre Whole (*Wm*)	Analogous to *Dm*, but applies only to unbroken blots.†	I. "Bat," "fox's head" VI. "Mud turtle"
5	Plus Detail (*D*+)	Two or more *D* areas are combined into one "good form" percept.*	II. (*D* 1's) "Bears fighting"
	Plus Whole (*W*+)	All *D* portions of a broken plot are combined into one "good form" percept.*†	II. "Two fellows at a bar toasting each other"
6	Plus-Plus Whole (*W*++)	An unbroken blot is perceptually articulated and reintegrated into a "good form" percent.†	IV. "A giant sitting on a stump"
	Plus-Plus Detail (*D*++)	A *D* area is articulated and reintegrated into a "good form" percept.	X. (*D* 8, left) "Guy riding a horse"

* Beck's tables (3,4) are used as a guide in scoring.

† Unbroken blots are I, IV, V, VI, and IX; broken blots are II, III, VII, VIII, and X.

Level six is characterized by the highest form of differentiation and hierarchic integration which is found only in mature perception.

In scoring, additional responses given during the inquiry were included. If *S* rejected a response on inquiry, it was excluded. If *S* improved a response on inquiry, he was given appropriate credit. The focus was on how well *S* could do rather than on how poorly. Beck's locations (3, 4) were used for *D* and *Dd* areas.

Siegel (23) found the mean percentage of agreement among three judges to be 93.9 per cent, using all of the above scoring categories with the exception of

perseveration and the distinction between *Dd*+ and *Dd*−. His findings indicated that the definitions are clear enough to be used by other Rorschach examiners. No further reliability studies were made at this time.

Proverbs test. Benjamin (9) was one of the first to apply proverbs in a systematic way to the study of schizophrenic thinking. As Werner's theories would suggest, the poorly differentiated schizophrenic and the child tend to take a literal or concrete approach to proverb interpretation. In administering the test, *S* was given an example and told to "interpret the proverb, to try to give a general meaning." Thirteen of the 14 proverbs were then given in the order originally used by Benjamin (9). Proverb 13 (an absurd phrase and not a proverb) was omitted. While Benjamin has never published a systematic scoring system for his test, he has described the types of thinking pathology exhibited through proverb interpretation (9). In developing the scoring system, both Benjamin's analysis and Werner's theories were taken into consideration. The following general scoring classifications were used[7]:

Abstract III: A correct generalized interpretation without detracting elements. (Weight of 6.) Example: "Brooding over past mistakes is futile."

Abstract II: A correct example with reference to human behavior; another proverb meaning the same thing; a response partly generalized, partly restricted to a specific example; a lower level generalization. (Weight of 5.) Example: "What's done is done."

Abstract I: A response tinged with the literal; a response which would be acceptable at Abstract II but for some minor inaccuracy, overstatement, or alternative explanation which is false. (Weight of 4.) Example: "*Don't cry over* something that's happened; can't be helped" (literal tinge).

Vague Response: An attempt at interpretation which is on the right track, but is left too vague to be adequate, or fails to account for part of the proverb. (Weight of 3.) Example: "It's too late, in other words."

False Interpretation: The interpretation is very inaccurate, yet an attempt was made to interpret. The error is usually due to faulty desymbolization or faulty generalization. (Weight of 2.) Example: "Don't let defeat stop you."

General Literal: The interpretation is literal in effect, though stated in general terms. At first glance these responses do not appear to be literal, but they can only be understood as stemming from a literal interpretation. (Weight of 2.) Example: "What is wasted is wasted; no tears."

Literal: The proverb is interpreted literally. (Weight of 2.) Example: "Don't cry when you spill some milk."

Absurd: The response indicates a failure to interpret and/or is logically absurd in terms of the task at hand. These responses are usually based on concrete associations to some aspects of the stimulus. (Weight of 1.) Example: "The milk is on the floor and the horses will drink the water."

Literal-abstract: A response which gives both a literal and an abstract interpretation. The tendency to be drawn into a literal interpretation is strong, but the subject is able to counteract it. (Weight at best abstract level reached.) Example: "Call the cat to lick it up, or what is happened has happened; let's look ahead."

To determine rater reliability, 25 cases were rescored by a fourth-year clinical psychology student. This second rater used the general definitions given above, plus examples drawn from the cases not rescored. The two-rater reliability was found to be .98. Odd-even reliability for 50 cases was found to be .83. It would appear that the scoring system is reliable and that the different parts of the Proverbs test define a similar function.

Because of a high relationship between ability to interpret proverbs and verbal intelligence in normals (18), it was assumed that a more sensitive index of thinking disturbance would be a discrepancy score based on the standard-score difference between a vocabulary estimate of verbal intelligence and the Proverbs score.[8, 9]

[7] The reader is referred to the original work (6) for more detailed scoring instructions and for examples to be used as a guide in scoring. The examples given here are taken from the second proverb, "Don't cry over spilt milk." It should be noted that this scoring system does not have the empirical foundation that Friedman's and Siegel's studies gave to the Rorschach. Genetic studies might help to improve this system.

[8] The adequacy of this assumption is tested later. The usual argument is that the degree of disturbance in mental organization is best measured by the loss relative to the individual's own potential rather than loss relative to an average potential. Babcock (1), Shipley (22), and Simmins (24) have all supported the pragmatic value of this contention. However, this would imply that mental maturity is a relative concept, whereas in scoring the Rorschach, maturity levels are generally defined in terms of absolute standards. This dilemma seems to have arisen because of differential knowledge in the areas of perceptual organization and conceptual organization. In the former case, there are no ready tests and data suggesting how well a person might do under optimal conditions (*i.e.*, without psychosis), so judgments of severity are made on the basis of an average expectation (absolute standards). In the latter case, there are ways of estimating one's potential, and this information can be used to gain a more refined prediction. The basic source of difficulty seems to lie in the equation of immaturity with severity of illness. Perhaps immaturity-maturity might better be taken as the reference scale and severity of illness defined as the difference on this scale between potential and present levels of functioning. This problem needs further thought and study.

[9] The actual measure of vocabulary level used was the last word right before six consecutive failures. This measure correlates .907 with the usual way of scoring vocabulary, but seems to give a better estimate of potential verbal intelligence for the acutely disturbed and chronically withdrawn, who show greater inconsistency in response to any verbal test.

Evaluation of Process-Reaction from Case Histories

The method of quantifying the social history data so as to reflect the process-reaction dimension was adapted from Wittman (26, 27). Wittman has made repeated studies of the prognostic value of case history data and formalized her findings as the Elgin Prognostic Scale. The scale consists of 20 variables, weighted according to prognostic significance. Included in the subscales are evaluations of prepsychotic personality, nature of onset, and typicality of the psychosis relative to Kraepelin's definition. These are the same factors on which the process-reaction distinction was made, so that Wittman's scale is an adequate summary of what has typically been included in the process-reaction distinction.

Three primary sources of data were used in rating the Elgin scale: (*a*) direct interview with the patient, (*b*) anamnesis, and (*c*) psychiatric history obtained by the examining psychiatrist. Using her scale with similar sources of information, Wittman found a two-rater reliablity of .87 for 61 cases (26). In several studies of the prognostic validity of the scale, Wittman (26, 27, 28) found the scale to predict outcome, with or without shock therapy, in 80 to 85 per cent of the cases. These data indicate that this is a moderately reliable instrument for assessing the process-reaction dimension.

Wittman has defined only the end points of each subscale. To add precision to the ratings, definitions were developed for intermediate points.[10] The added definitions adhered closely to Wittman's original intent as to subscale meaning. Though a check of the reliability of the modified scale was not possible, it is doubtful if it would be less than that reported by Wittman.

PREDICTIONS

The following specifications of the hypothesis were made: Those schizophrenics with more process-like case histories (higher scores on the Elgin scale) receive significantly lower mean-genetic-level scores on the Rorschach and significantly lower Proverbs-minus-vocabulary discrepancy scores than those schizophrenics with more reactive-like case histories. The critical acceptance level was set at .01, using one-tailed tests of significance.

RESULTS

The distribution of scores on the Elgin scale were first normalized by use of McCall's *T*-scored method (11). Analysis showed a significant difference (.02 level) in the means for the men and women on the Elgin scale. The mean for the men was 52.9 and for the women 46.2. This difference suggests that on the average the women in this sample were less severe cases. A possible selective factor is the care given a large portion of the less severe males by the Veterans Administration hospitals in the San Francisco area. Because of this significant difference in means on the Elgin scale, the hypotheses were first examined separately by sex group. If the relationships did not differ significantly, the correlations were then averaged to obtain an estimate of the population r independent of the sex difference in means on the Elgin scale.

Main Results

The Rorschach hypothesis was confirmed for both sex groups. The Rorschach mean-genetic-level score and the Elgin Prognostic Scale correlated $-.599$ ($p < .01$) for the men, and $-.679$ ($p < .001$) for the women. Since these correlations did not differ significantly, an averaged r was computed and found to be $-.641$ ($p < .001$).[11] These results indicate that there is a meaningful relationship between the process-reaction dimension, as evaluated from case-history data, and disturbances of thought processes as measured by the Rorschach genetic-level scoring system.

The hypothesis concerning the Proverbs test was confirmed only for male schizophrenics. The correlation between the Proverbs discrepancy score and the Elgin scale was $-.682$ ($p < .001$) for men, and .048 (obviously insignificant) for the women. The difference between these two correlations is significant beyond the .01 level. These results indicate that the Proverbs-vocabulary discrepancy score is significantly related to the process-reaction dimension for men, but not for women.

Further Analyses

In making further analyses, certain interrelations of the data were explored. First, it was found that the Rorschach and Proverbs-vocabulary discrepancy scores measured a similar function for the men ($r = .599$; $p < .01$) but not for the women ($r = -.210$). Second, the adequacy of the assumptions about vocabulary corrections was tested, using a multiple correlation technique. No predictive gain was found when vocabulary was used in the re-

[10] See Appendix I in Becker (6).

[11] Fisher's z transformation was used to compute the averaged r (19).

gression equation to predict Elgin-scale placement from the Rorschach. This result is consistent with the initial assumption.[12] A similar analysis for the Proverbs test (meaningful only for the men) showed an increase in correlation from −.603 for Proverbs alone to .690 (corrected for "shrinkage") when both Proverbs and vocabulary were used to predict the Elgin scale.[13] The multiple R gives about the same result as was reported earlier for discrepancy scores (−.682) and supports the value of a vocabulary correction. Finally, the effectiveness of the best linear combination of the Rorschach, Proverbs, and vocabulary scores in predicting the Elgin scale was investigated. For the men, the multiple R was .768 after correction for "shrinkage." All three tests contributed substantially to the prediction with vocabulary acting as a suppressor variable. For the women, the multiple R showed no increase over a direct prediction from the Rorschach, since the Proverbs test carries no weight for this group.

DISCUSSION

The results in part support the central hypothesis and demonstrate the value of a levels-of-personality interpretation of the process-reaction construct. A number of questions, however, were raised by the findings. Most difficult to understand is the sex difference in the meaning of the Proverbs test. Elaborate statistical analyses of the data failed to clarify this situation. A possible clue may lie in the fact that predictions were confirmed as long as the focus was on perceptual organization (Rorschach) but failed to be confirmed for women with the shift to conceptual organization (Proverbs). Possibly conceptual functioning in schizophrenic women is more easily disrupted than perceptual functioning by momentary emotional factors independent of severity of illness. However one interprets these results, if supported by future research, they imply a need either to control for the "interfering factor" or to develop different types of instruments for the evaluation of degree of pathology in male and female schizophrenics.

Two further implications should be mentioned. First, the Rorschach findings may be interpreted as evidence for the validity of the Rorschach in reflecting levels of personality organization. In addition, a theoretical framework (seldom found in the Rorschach literature) is offered for why the Rorschach should be able to perform such a task. Second, the coherence of the findings suggests the feasibility of developing a diagnostic system similar to the stanine system used by the Air Force. On the basis of history and test data, schizophrenics could be classified as Level One's, Level Two's, etc., and the diagnosis would have genuinely prognostic value. In addition, such a classification would provide a sounder basis for distinguishing subgroups for etiological research, an important control variable for research on the effectiveness of therapy, and a frame of reference for evaluating the course of the disturbance over time.

SUMMARY

The Elgin Prognostic Scale was used to evaluate the case records of 24 male and 27 female schizophrenics in terms of a process-reaction continuum on the assumption of its equivalence to a measure of level of personality organization. Werner's developmental theories and certain empirical studies were used to score Rorschach and Benjamin Proverbs responses in terms of levels of differentiation and integration. It was then predicted that more process-like schizophrenics obtain lower Rorschach genetic-level scores and lower Proverbs-minus-vocabulary scores than more reactive-like schizophrenics. Because of a significant difference between men and women on the Elgin scale, the hypotheses were examined separately for the sex groups.

The Rorschach hypothesis was confirmed for both sex groups. The r between the Rorschach genetic-level score and the Elgin Prognostic Scale was −.599 ($p < .01$) for the men and −.679 ($p < .001$) for the women. The hypothesis for the Proverbs test was confirmed only for male schizophrenics. The r between the Proverbs discrepancy score and the Elgin scale was −.682 ($p < .001$) for the men and .048 for the women. No adequate explanation for this sex difference was found.

[12] A correction based on some measure of potential level of perceptual organization might add to the predictive efficiency of the Rorschach test.

[13] All multiple Rs are positive in sign. The implied relationship in this case is the same as if the multiple R had a negative sign.

The general conclusion is drawn that the results in part support the central hypothesis. There is evidence for a measurable dimension of regressive and immature thinking which is related to the process-reaction dimension. In addition, the value of a levels-of-personality-organization interpretation of the process-reaction dimension is supported.

REFERENCES

1. BABCOCK, HARRIET. *Dementia praecox, a psychological study.* Lancaster, Pa.: Science Press, 1933.
2. BALDWIN, A. L. *Behavior and development in childhood.* New York: Dryden, 1955.
3. BECK, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
4. BECK, S. J. *Rorschach's test.* Vol. I. *Basic processes.* (2nd Ed.) New York: Grune & Stratton, 1949.
5. BECKER, W. C., & MCFARLAND, R. L. A lobotomy prognosis scale. *J. consult. Psychol.*, 1955, **19,** 157–162.
6. BECKER, W. C. The relation of severity of thinking disorder to the process-reactive concept of schizophrenia. Unpublished doctor's dissertation, Stanford Univer., 1955.
7. BELLAK, L. *Dementia praecox.* New York: Grune & Stratton, 1948.
8. BELLAK, L. A multiple-factor psychosomatic theory of schizophrenia. *Psychiat. quart.*, 1949, **23,** 738–755.
9. BENJAMIN, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. of California Press, 1946. Pp. 66–71.
10. FRIEDMAN, H. Perceptual regression in schizophrenia. An hypothesis suggested by use of the Rorschach test. *J. proj. Tech.*, 1953, **17,** 171–185.
11. GUILFORD, J. P. *Fundamental statistics in psychology and education.* (2nd Ed.) New York: McGraw-Hill, 1950.
12. HEMMENDINGER, L. Perceptual organization and development as reflected in the structure of Rorschach test responses. *J. proj. Tech.*, 1953, **17,** 162–170.
13. HUNT, R. C., & APPEL, K. E. Prognosis in psychoses lying midway between schizophrenia and manic-depressive psychoses. *Amer. J. Psychiat.*, 1936, **93,** 313–339.
14. KANT, O. Differential diagnosis of schizophrenia in the light of concepts of personality stratification. *Amer. J. Psychiat.*, 1940, **97,** 342–357.
15. KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17,** 157–162.
16. LANGFELDT, G. Prognosis in schizophrenia and factors influencing course of disease: catamnestic study, including individual re-examinations in 1936 with some considerations regarding diagnosis, pathogenesis and therapy. *Acta Psychiat. Neurol.*, 1937, Suppl. 13. Pp. 1–228.
17. LEWIN, K. *A dynamic theory of personality.* D. K. Adams & K. E Zener (Trans.). New York: McGraw-Hill, 1935.
18. MCNEMAR, Q. *The revision of the Stanford-Binet scale.* New York: Houghton Mifflin, 1942.
19. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.
20. PHILLIPS, L., & SMITH, J. G. *Rorschach interpretation: advanced technique.* New York: Grune & Stratton, 1953.
21. PIOTROWSKI, Z. A. On the Rorschach method and its application in organic disturbances of the central nervous system. *Rorschach Res. Exch.*, 1936–1937, **1,** 65–77.
22. SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.*, 1940, **9,** 371–377.
23. SIEGEL, E. L. Genetic parallels of perceptual structuralization in paranoid schizophrenia: an analysis by means of the Rorschach technique. *J. proj. Tech.*, 1953, **17,** 151–161.
24. SIMMINS, C. Studies in experimental psychiatry: IV. Deterioration of 'G' in psychotic patients. *J. ment. Sci.*, 1933, **79,** 704–734.
25. STOTSKY, B. A. A comparison of remitting and nonremitting schizophrenics on psychological tests. *J. abnorm. soc. Psychol.*, 1952, **47,** 489–496.
26. WITTMAN, PHYLLIS. Scale for measuring prognosis in schizophrenic patients. *Elgin State Hosp. Papers*, 1941, **4,** 20–33.
27. WITTMAN, PHYLLIS. Follow-up on Elgin prognosis scale results. *Illinois psychiat. J.*, 1944, **4,** 56–59.
28. WITTMAN, PHYLLIS, & STEINBERG, D. L. Follow-up of objective evaluation. *Elgin State Hosp. Papers*, 1944, **5,** 216–227.
29. WERNER, H. *The comparative psychology of mental development.* (Rev. Ed.) Chicago: Follet, 1948.

Received November 21, 1955.

ASSESSMENT OF ABSTRACT BEHAVIOR IN A NON-WESTERN CULTURE

GUSTAV JAHODA

University College of the Gold Coast

CROSS-CULTURAL studies of test intelligence, and critiques of such studies abound in the literature and have been summarized by Anastasi and Foley (1), but abstract behavior or its alleged absence more often forms the subject of impressionistic judgments. Carothers discusses some of these with special reference to Africans and does not himself dissent from the view that the African "is inapt for sound abstraction and logic" (9, p. 87). Another recent example of sweeping generalization in this field is to be found in a series of papers by Haward and Roland (13, 14, 15), reporting the results of administering the Goodenough Draw-a-Man Test to some inadequately specified samples of Nigerians. Apparently not fully aware of the limitations of this test when used in non-Western cultures (12), these authors write about "the concreteness of the Nigerian's mental approach, a concreteness which is so rigid as to produce schizophrenic signs" (13, p. 88).

Until the work of McConnell (20) no attempt appears to have been made to assess the capacity for abstract behavior of non-Western people by means of the tests specially devised and commonly used in Western countries for this purpose. It was not the aim of the present study to discover the extent to which the unflattering statements about the African's "concrete mentality" would seem to be confirmed or contradicted by such a measure as the Goldstein-Scheerer Cube Test (GSCT); but rather to examine the question whether a test of this kind is any more appropriate for making cross-cultural comparisons than various types of intelligence tests, some of which it closely resembles.

METHOD

Subjects

The work was carried out in three stages, with independent groups of adolescent boys. Except for a small minority from other parts of the West Coast they were Gold Coast Africans. Over two-thirds belonged to the Ga tribe, indigenous to the Accra region; the remainder were mainly the sons of Ewe or Akan migrants. In the absence of universally compulsory registration of births, ages are difficult to ascertain; they ranged from 13 to about 18 years, but the modal ages varied somewhat. All *S*s were attending middle schools and had received eight or nine years of elementary education.[1]

The first group, whose modal age was about 15 years, took the GSCT. The group consisted of 27 boys from the Accra urban area, representing the remnant of an original sample of 33,[2] which had been selected in the following way: on the basis of information supplied by the education authorities concerning geographical and social "catchment areas" two schools were chosen as appearing to offer jointly a fairly satisfactory cross section of the middle-school population. Within the schools the third or penultimate form was decided upon after balancing the considerations of adequate command of English vs. duration of availability for investigation. Although some work could be done with the complete forms, for the purpose of intensive study two subsamples of 18 and 15 boys respectively were drawn by means of a table of random numbers.

At the time of the GSCT the investigator had been in contact with these boys for nearly four months. The home of each one had been visited, so that it was possible to decide reliably whether their home backgrounds were literate or illiterate. To ascertain this is a less simple matter than might be imagined, and the mere knowledge of the father's educational level is insufficient; children of a literate father may spend their lives with an illiterate mother, perhaps the second wife in a polygynous family, or with other illiterate relatives.

The second sample consisted of ten complete forms,

[1] It would be desirable to relate the school population from which these samples were drawn to the entire age groups. Unfortunately no adequate statistical data about age distribution in the general population is available. An estimate has, however, been published (28) of the total number of children of middle-school age in the Gold Coast as a whole. The sex ratio being known (30) it is possible to compare this estimate with the figures of boys' middle-school enrollment (29). From this one can roughly assess the proportion of eligible boys who attended middle schools at the time of the investigation as just under a half (it may be of interest that the proportion was only a tenth in the case of girls). In the Accra region the rate of enrollment was probably substantially higher. In the writer's view the original estimate is subject to a very wide margin of error; the inferences based on it are therefore hardly more than informed guesses.

[2] Two went to secondary schools; two more failed and had to leave; one moved to another town, and one became delinquent and was expelled.

numbering over four hundred, from several urban schools. These were given Raven's Progressive Matrices Test (23) on three successive weeks, and 317 boys completed the whole series. Their modal age was approximately 16 years.[3]

While the investigator was working in a semirural school he came across the work of McConnell (20). It was possible to replicate his study with one complete form of 32 boys, whose modal age was again about 15 years.

The medium of instruction in middle schools is English, but practically all the boys speak only the vernacular at home. Great care was therefore taken to ensure that test instructions were clearly understood.

Tests Employed

Goldstein-Scheerer Cube Test. The terms "abstract" and "concrete" are sometimes used as if they were descriptive of two contrasting types of mentality. In fact, as with nearly all psychological variables, one is dealing with a continuum. In their monograph (11, p. 48 ff.) Goldstein and Scheerer claim that the GSCT can serve to differentiate between abstract and concrete types of normals, and it thus appeared a suitable tool for exploring the problem. Further advantages are the quantitative scoring system and rough norms devised by Boyd (8), which facilitate comparison.

The question of whether the Cube Test would be valid in other cultures was not explicitly faced by Goldstein and Scheerer. It must be remembered that they are primarily concerned with the diagnosis of pathological impairment, and "normal" in their usage probably has to be understood in this sense. They do state quite clearly, however, that all "normals" are equipped to solve the Cube tests, at any rate with the aids provided.[4] It would seem to follow that "normal" people in other cultures should also be able to perfom successfully.

Some of Goldstein and Scheerer's underlying assumptions will be examined later, but it might be mentioned at this stage that Boyd (8) was able to demonstrate a relationship between subaverage Wechsler-Bellevue IQ level and GSCT performance; only *S*s with an IQ of 100 or over obtained, as a group, a perfect Cube Test score. This finding indicates the existence of common factors determining performance on tests of intelligence and of abstract behavior, and suggests that the latter, like the former, may not be culture-free.

In adminstering the GSCT the procedure laid down by the authors (11) was followed in every detail.

Raven's Progressive Matrices. This test requires the discovery of abstract relationships among designs of increasing complexity. It consists of symmetrical patterns, in which one element is missing. Below each pattern a set of shapes is printed, from which the one that will lead to correct completion has to be selected. This test was used during World War II with African recruits, and obtained a factor loading closely similar to that of the Block Design Test (25, p. 106), on which the GSCT is based.

The way the Progressive Matrices were applied owes its inspiration to Ombredane (22), who suggested that European tests may be more effective with Africans when administered more than once, thereby functioning as tests of educability rather than intelligence. Hence it was given on three successive weeks, with a 20-minute time limit. Otherwise Raven's instructions (23) were adhered to.

Kohs Block Design Test. The Wechsler-Bellevue version of this test ("with certain modifications in presentation suggested by Goldstein and Scheerer") had been applied by McConnell (20) to Tepehuan Indian villagers. It was administered in the manner outlined by him, and his five scoring categories were adopted.

RESULTS

Goldstein-Scheerer Cube Test

Qualitative aspects of performance. In attempting the first design most *S*s clung persistently to a single block. When reminded that all four were to be used they tended to build towers like some of the "normals" reported by Goldstein and Scheerer. On the next step being presented they would reproduce the pattern correctly, but with the return to the original design *S*s were apt to revert to the single block. The reason emerged quite clearly from their behavior and comments: most of them did not at first perceive the various steps as pertaining to the same task; when the original was shown again they responded once more in what they felt to be the correct way. The whole procedure was quite strange to them, so they were not surprised or annoyed on being called upon to do the same thing several times over again.[5] With the next few designs, where more than one color was involved, the function of the several steps came to be appreciated.

The most widespread deficiency of the boys'

[3] This part of the investigation was carried out jointly with Mr. A. Taylor of the Institute of Education, and the writer is indebted for permission to use some of the material for this paper.

[4] "Should he (i.e. the 'normal') have failed on the standard design, he invariably will succeed on any one of its simplifications and also on the same standard design if presented again" (11, p. 55).

"There is definite indication of impairment of abstract behavior if a subject . . . fails again on the standard design" (11, p. 56).

[5] Similar behavior was observed when the writer used mental mazes with African *S*s. Willing to oblige the experimenter and not in a hurry to get somewhere, they may respond with the same letter for many minutes on end. One highly intelligent student (Otis Gamma IQ of 114) took over three hours to get to the end of a maze which English *S*s of much lower IQ completed in less than 30 minutes.

performance, which caused them to lose more points than any other type of failure, does not seem to have been considered at all by Goldstein and Scheerer. In their reproductions *S*s paid very scant attention to the spatial arrangement of the blocks on the table; designs were usually left at whatever angle they happened to be completed, so that there was nearly always some rotation. The *S*s on the whole did not regard the figure-ground relationship as part of their task. This came out most clearly in cases where the multiple-choice stage was reached; when three correct models were rotated 0, 45, and 90 degrees, the answer to inquiry was almost invariably to the effect that all three were equally right.

Ignoring spatial position is something that can be fairly easily corrected by specific and emphatic instructions, as subsequent check trials with other *S*s proved. Another, more fundamental aspect of the performance of the majority of the *S*s was not so easily subject to improvement. It will be best to illustrate this by the actual record of the behavior of one boy.

At the outset he looked at the printed design for about 12 seconds, then turned the blocks and put them together in various ways without even once glancing at the design until 1 minute 40 seconds; during this time he once achieved the correct pattern and destroyed it again; after another 10 seconds there followed a further brief glance, and then nothing until 2 minutes 35 seconds when he examined the design for about 8 seconds; a last glance occurred just before the time was up.

It is clear from this that the procedure was not a systematic block-by-block matching, with a continuous "feedback," but rather the reproduction of an infrequently renewed memory image that was often faulty. With the designs of greater complexity a more analytical approach was frequently though not always adopted. Check trials were carried out later in an attempt to induce *S*s to use the block-by-block matching method from the outset, but the success was only partial. This observation of course does not prove that it could not be done; on the contrary, the writer gained the impression that it would be possible with a more prolonged training. It is therefore doubtful whether the behavior described is the same as Goldstein and Scheerer's "concrete dependence upon the impression of the figure, and the inability to break up the figure ideationally into single squares" (11, p. 39).

Apart from the causes of failure outlined above, several of Goldstein and Scheerer's categories were also applicable occasionally. The ones below, relating to the presented model, are listed in decreasing order of importance (numbers refer to the monograph):

3. Concrete dependence upon one global impression of the design—without articulation of parts;
4. Concrete dependence upon the "impressiveness" of certain color aspects;
1. Concrete dependence upon size.

On the whole, however, these seem to have played a relatively minor part. Failure responses to the manipulative block material occurred sometimes at the beginning, but were usually eliminated after a little handling of the blocks.

Numerical results. The mean score for the whole group was 77.6, with an *SD* of 20.3, and the range 34–115; as the possible maximum score is 120, this means that not one *S* achieved a faultless performance.

When the group is divided according to home background, an important contrast emerges. There were 17 boys from literate homes, whose mean score was 84.9; the remaining 10 from illiterate homes averaged only 65.1, a difference significant at the .02 level. The influence of environmental factors is thus clearly evident.

It would seem, therefore, that the low level of performance does not necessarily involve a lack of abstract ability, but perhaps only an inadequate development of perceptual skills. If such is the case, one would expect an improvement after a certain amount of practice with the handling of geometrical designs. On the other hand, if Africans are "incapable of sound abstraction," practice should make no difference, for Goldstein and Scheerer state that people with "impairment of abstract behavior are incapable of learning, or of acquiring the abstract attitude by learning" (11, p. 56).

Raven's Progressive Matrices

It appeared advisable to test the conclusion just outlined with a larger sample. The GSCT

TABLE 1
CUMULATIVE PERCENTAGES OF PROGRESSIVE MATRICES SCORES
(Maximum score is 60)

Group	Scoring Levels								
	50+	45+	40+	35+	30+	25+	20+	15+	0+
2,010 British farm-workers aged 17–19	1	8	25	48	69	85	94	98	100
317 African school-boys (first week)	.3	1	4	17	31	46	61	74	100
317 African school-boys (third week)	.6	9	30	46	58	64	71	82	100

does not easily lend itself to such use, being extremely time-consuming to administer; hence the use of the Progressive Matrices Test, which taps somewhat similar abilities. As mentioned earlier the test was given on three successive weeks, with a 20-minute time limit. Table 1 shows the distribution of scores, in comparison with that of a sample of British farm workers of a somewhat higher age group. The British norms were obtained from Vernon (24), and farm workers chosen as being the group with, presumably, the least dissimilar background.

It will be seen that the scores in the third week approximate fairly closely those of the British farmworkers. The mean rise was 6.5 points or 28 per cent, a highly significant change with a critical ratio of 16.28; and contrary to the findings of Vernon and Parry (26) on a retest of British *Ss*, the greatest improvement took place in the middle ranges.

With a correlation of .82 between the two sets of scores the changes exhibited considerable variations, ranging from minus 24 to plus 32; uncontrolled motivational factors may have been at least partly responsible. Preliminary trials with untimed tests had shown that some of the boys became bored after the second test. That they probably lost interest to some extent even when there was a time limit is suggested by the fact that over one-quarter of the *Ss* declined between the second and third test, sometimes quite markedly; the extreme case of one boy may be quoted who scored 28, 35 and 4 over the series. The results would have been far more spectacular if one had selected the best scores from all three tests, but such a procedure would have capitalized on chance successes. It may be stated with confidence, however, that the rates of improvement reported are on the conservative side.

These results support the view that increasing familiarity with the materials among which abstract relationships are to be discovered leads to a significant improvement on the part of African *Ss*.

Kohs Block Design Test

The proportion who came from literate homes was far smaller in this group of *Ss* than among those who had done the GSCT; even so the extent of literacy may be exaggerated, for it was necessary in this case to rely on indirect information obtained by questioning the boys. Yet the comparative results presented in Table 2 fail to indicate that the *Ss* are markedly inferior to the Tepehuan, of whom McConnell said that they function at an abstract level "almost as well as the adult population on which Wechsler standardized his intelligence test."

Thus it would seem that the African urban sample, as judged by the GSCT results, tend to be very concrete in their behavior, but a scholastically poorer semirural group could be said to be nearly as abstract as a representative Western group on the basis of their performance on the modified version of Kohs Blocks. This apparent contradiction calls for some explanation. There are three main reasons why performance on this test was relatively easier, and the handicap of the *Ss* from illiterate homes considerably reduced:

1. The absence of single-color designs (except among the sample items), which elimi-

TABLE 2
RANK SCORES ON KOHS BLOCKS

Rank	Adult Tepehuan Indians*			Adolescent African Schoolboys		
	Sex			Background		
	Men	Women	Total	Literate	Illiterate	Total
I	0	0	0	0	0	0
II	2	2	4	0	1	1
III	3	4	7	4	8	12
IV	7	5	12	10	8	18
V	1	0	1	0	1	1
Total	13	11	24	14	18	32

* From McConnell (20).

nated the occasion for what Goldstein and Scheerer called "concrete dependence upon size," actual or apparent.

2. The absence of diagonal designs, which meant that rotations could be ignored, as there were no specific instructions to the contrary.

3. Most important of all were the two sample designs, which enabled the experimenter to explain the task in detail to the *S*s, and to ensure that they grasped the basic mode of approach before they had to start working independently.

Some of the characteristic failure responses described earlier recurred, but owing to the different conditions of administration they had less effect on the final scores. Unlike the Tepehuan sample none of the boys would have received additional credits for rapid working if Wechsler's scoring instructions (27, p. 184) had been followed.[6] On the other hand only one single *S* out of the total group could be classified, according to McConnell's scheme, as being more concrete than abstract in his performance.

DISCUSSION

The series of investigations reported indicate that the problems inherent in any attempt to assess abstract behavior in different cultures are closely analogous to those encountered when one is trying to compare levels of intelligence. Neither can operate in a vacuum, and the media through which they are tested are not equally familiar to members of different cultures. It is one of the weaknesses of Goldstein and Scheerer that they do not seem to recognize this, for they write (11, p. 29): "... every normal individual—even a deaf-mute child—possesses these basic categories of like shape and color." As far as Gold Coast Africans are concerned, this is certainly not true of shape. Hebb (16) has shown that the perception of shapes is the result of a lengthy learning process. In a cultural milieu where geometrical patterns form part of a person's early experience only to a very limited extent or not at all, as is the case with most rural and many urban Africans, the discrimination of complex shapes is likely to be a most exacting task.

Various investigations in different parts of Africa have accumulated data showing the difficulties Africans experience in manipulating spatial relations and perceiving complex shapes (2, 5, 6, 10, 21). While the possibility of some contributory genetic factor cannot be excluded a priori, the available evidence as marshalled, for instance, by Biesheuvel (4) suggests that the absence of adequate environmental support for such abilities is mainly or perhaps entirely responsible. The findings with Raven's Progressive Matrices are in line with this view.

It follows that neither the GSCT nor McConnell's adaptation of the Wechsler-Bellevue version of Kohs Block Test are suitable instruments for measuring abstract behavior in Africans, and presumably also in members of other non-Western cultures. This conclusion emerges sharply from the paradoxical results obtained in the present study with these tests.

To say this, of course, by no means implies agreement with those who hold that Africans are by nature concrete in their mentation. On the contrary, many of the African indigenous games (7, p. 110), and in the Gold Coast particularly Wari (3), require abstract ability of a high order. Thus the present writer is in agreement with part of McConnell's conclusion, though not with the method by which it was reached.

In this connection it may be observed that McConnell's belief that he discovered or proved something about "primitive thought" is based on a confusion. The misconception is best illustrated by his reference to Lévy-Bruhl, whom he describes as "contending that the thought patterns of *l'homme sauvage* are qualitatively different from those of 'modern man'...." There is an ambiguity here in the use of the term "thought pattern," which may relate to the content, a social fact, or to the process, a psychological one. In his later formulation[7] Lévy-Bruhl made

[6] The mean times taken for completion, converted into percentages of the maxima permitted by Wechsler, ranged from 85 per cent for the first design to 345 per cent for the last one.

[7] One has to admit that in his major publications Lévy-Bruhl's mode of expression was sometimes ambiguous, and this has led to a widespread misunderstanding of his position. In fact, as the quotation below from a work published in 1910 indicates, he never

it quite clear that he meant the former (18, 19), and therefore no performance on a test of abstraction of the present kind, however superior, is capable of refuting him. The investigation of "primitive man's mode of thinking" would need a very different type of approach.

SUMMARY

The aim of the study was to investigate whether tests of abstract behavior like the Goldstein-Scheerer Cube Test can be validly applied to non-Western people. All *S*s were adolescent boys from schools in or near Accra.

The Goldstein-Scheerer Cube Test was administered to 27 *S*s. The main types of failure responses that occurred are described. It was shown that boys from literate homes performed significantly better than those from illiterate ones, evidence that Cube Test results are not unaffected by environmental influences.

In order to check this finding with a larger sample, a group test involving similar abilities was given on three successive weeks, a highly significant practice effect being obtained.

A replication of McConnell's work with the Tepehuan, using a method derived from Goldstein and Scheerer, yielded results inconsistent with those from the Cube Test. The implications of this finding were discussed, and Goldstein and Scheerer's assumptions about perception criticized in the light of research in Africa. The conclusion emerged that tests of abstract ability are no more "culture-free" than tests of intelligence.

seems to have held the view that the basic cognitive processes of "primitives" differ from those of people in more advanced societies, except in so far as they are subject to social influences.

"In this way, the exigencies of logical thought are elicited, established and then strengthened in each individual mind by the uninterrupted pressure of the social milieu, by means of language itself and through what is transmitted in the forms of language. This constitutes a heritage of which no one is deprived in our society and which no one can even have the thought of rejecting . . .

"Quite other are the conditions where the prelogical mentality obtains. No doubt it is also socially transmitted, through the intermediary of language and of concepts without which it could not find expression . . . But these concepts differ from ours and, consequently, the mental operations differ also" (17, pp. 113–4) (Writer's translation).

REFERENCES

1. ANASTASI, ANNE, & FOLEY, J. P. *Differential psychology*. New York: Macmillan, 1949.
2. BARRÉ, MLLE. R. Psychologie. In L. Pales (Ed.), *Organisme d'enquête pour l'étude anthropologique —rapport No. 3*. Dakar: Gouvernement Général de l'Afrique Occidentale Francaise, 1947. Pp. 99–113.
3. BENNETT, G. T. Wari. In R. S. Rattray, *Religion and art in Ashanti*. London: Oxford Univer. Press, 1927. Pp. 382–390.
4. BIESHEUVEL, S. *African intelligence*. Johannesburg: South African Institute of Race Relations, 1943.
5. BIESHEUVEL, S. Psychological tests and their application to non-European peoples. In G. B. Jeffery (Ed.), *The yearbook of education 1949*. London: Evans, 1949. Pp. 87–117.
6. BIESHEUVEL, S. The study of African ability. Part I: The intellectual potentialities of Africans. *African Stud.*, 1952, **11**, 45–57.
7. BIESHEUVEL, S. The study of African ability. Part II: A study of some research problems. *African Stud.*, 1952, **11**, 105–117.
8. BOYD, F. A provisional quantitative scoring with preliminary norms for the Goldstein-Scheerer Cube Test. *J. clin. Psychol.*, 1949, **5**, 148–153.
9. CAROTHERS, J. C. *The African mind in health and disease*. Geneva: World Health Organization, 1953.
10. GENEVAY, M. J. La formation sociale et professionelle des noirs. *Rev. Mens. Comité National Organ. Francaise*, 1949, **23**, 13–16.
11. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
12. HAVIGHURST, R. J., GUNTHER, MINNA K., & PRATT, INEZ E. Environment and the Draw-A-Man test: the performance of Indian children. *J. abnorm. soc. Psychol.*, 1946, **41**, 50–63.
13. HAWARD, L. C. R., & ROLAND, W. A. Some intercultural differences on the Draw-A-Man test: Goodenough scores. *Man*, 1954, **54**, 86–88.
14. HAWARD, L. C. R., & ROLAND, W. A. Some intercultural differences on the Draw-A-Man test: Part II, Machover scores. *Man*, 1955, **55**, 27–29.
15. HAWARD, L. C. R., & ROLAND, W. A. Some intercultural differences on the Draw-A-Man test: Part III, conclusion. *Man*, 1955, **55**, 40–42.
16. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
17. LÉVY-BRUHL, L. *Les fonctions mentales dans les sociétés inférieures*. (9th Ed.) Paris: Presses Universitaires de France, 1951.
18. LÉVY-BRUHL, L. *Les carnets de* . . . Paris: Presses Universitaires de France, 1949.
19. LÉVY-BRUHL, L. A letter to E. E. Evans-Pritchard. *Brit. J. Sociol.*, 1952, **3**, 117–123.
20. MCCONNELL, J. Abstract behavior among the Tepehuan. *J. abnorm. soc. Psychol.*, 1954, **49**, 109–110.
21. NISSEN, H. W., MACHOVER, S. & KINDER, E. F. A study of performance tests given to a group

of native African negro children. *Brit. J. Psychol.*, 1935, **25**, 308–355.

22. OMBREDANE, A. Principes pour une étude psychologique des noirs du Congo Belge. In H. Piéron, A. Fessard, & P. Fraisse (Eds.), *Année psychol.* Paris: Presses Universitaires de France, 1951. Pp. 521–547.
23. RAVEN, J. C. *Guide to using Progressive Matrices.* London: Lewis, 1948.
24. VERNON, P. E. Occupational norms for the 20-minute Progressive Matrices Test. *Occup. Psychol.*, 1949, **23**, 58–59.
25. VERNON, P. E. *The structure of human abilities.* London: Methuen, 1950.
26. VERNON, P. E., & PARRY, J. B. *Personnel selection in the British forces.* London: Univer. of London Press, 1949.
27. WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.
28. *Accelerated development plan for education, 1951.* Gold Coast Government. Accra: Government Printing Department, 1951.
29. *Annual report of the education department for the year 1952.* Gold Coast Government. Accra: Government Printing Department, 1954.
30. *Census of population 1948.* Gold Coast Government. Accra: Government Printing Department, 1950.

Received November 28, 1955.

THE MANIPULATION OF DOMINANCE IN MONKEYS WITH CONDITIONED FEAR

J. V. MURPHY AND R. E. MILLER[1, 2]

Department of Clinical Science, School of Medicine, University of Pittsburgh

CONDITIONING as a methodological tool achieved a major advance with the advent of instrumental conditioning in that the response restrictions imposed in the classical conditioning experiments of Pavlov were largely eliminated. That is, the subject was frequently free-ranging in the experimental apparatus instead of being securely fastened in a conditioning stock. This modification sacrificed some degree of control over the conditioned response but appreciably reduced the artificial nature of the learning process. In an attempt to extend this principle, a technique for interanimal conditioning was devised (8) which relinquishes the rigid experimental control over both the stimulus and the response. With this procedure the conditioned stimulus is a monkey which is free-ranging in a relatively large compartment. Thus, the stimulus animal may be moving about, grooming himself, or performing any one of many possible behaviors during presentation to the conditioning subject. Moreover, the conditioning subject is also permitted to move freely in a relatively large compartment. In order to acquire a CR the subject, from any position in the compartment, must identify the stimulus animal regardless of its position or behavior. One might anticipate that learning in such a situation would generalize to a greater extent to other situations since varying aspects of both the stimulus and the response are incorporated in the learning process.

Mowrer (6) has advanced an interpretation of avoidance learning which utilizes the intervening variables of fear and fear reduction. Applying this theoretical interpretation to interanimal conditioning, the sight of the stimulus monkey accompanied by a painful electric shock results in an association between the stimulus animal and the pain and fear evoked by the shock. On subsequent trials, the sight of the stimulus animal gives rise to fear in the animal undergoing conditioning. Performing the arbitrarily designated CR removes the CS (sight of the stimulus animal) and the conditioning subject is theoretically reinforced by fear reduction. A major difficulty with the fear-reduction interpretation is that the inferential evidence for the presence of fear and fear reduction is inextricably related to the conditioned response.

One advantage of the social avoidance procedure is that behavioral modifications are not necessarily restricted to the particular response or situation used to establish conditioning. If the sight of the stimulus animal gives rise to fear in the conditioned animal, the interactions of the two monkeys should be modified in a predictable fashion outside of the conditioning apparatus. Such a modification would provide independent evidence for acquired fear in avoidance learning. The dominance relationship would appear to be ideally suited for an independent assessment of "fearful" behavior. Maslow (3) has described the dominance behaviors of the Old World monkeys, including rhesus, as:". . . rough, brutal and aggressive, it is of the nature of a powerful, persistent, selfish urge that expresses itself in ferocious bullying, fighting, and sexual aggression. The subordinate animal is usually afraid of the dominant animal, and is frequently completely terror-stricken and cowed" (3, p. 316). If fear is induced in a dominant monkey by the sight of a submissive partner which has been used as a CS for avoidance behavior, one would predict that the dominance relationships would be altered after the completion of conditioning.

Specifically, the present experiment was designed as an attempt to verify and extend the observation (5) that the dominance hierarchy within a group of monkeys is subject to experimental manipulation with the interanimal conditioning technique. That is, can the dominant member of a pair of monkeys be made to behave as a submissive animal follow-

[1] This investigation was supported by a research grant (M-487C) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service; and in part by the Foundations' Fund for Research in Psychiatry.

[2] The authors are indebted to Dr. I. Arthur Mirsky for suggestions and criticisms.

ing avoidance conditioning in which the submissive member of the pair is the conditioned stimulus? An incidental aim of the present study was to obtain evidence on the presence of acquired fear in avoidance learning which is independent of the conditioning situation.

METHOD

Subjects. Ten experimentally naive male rhesus monkeys were used as *Ss*. The weight-range in the group was 9½ to 18 lbs. when the study was undertaken and 12 to 25 lbs. at the finish.

Apparatus. The conditioning apparatus (8, Fig. 1) consisted of a large rectangular box divided into two compartments by a specially constructed one-way vision screen which made it impossible to view the stimulus compartment from the conditioning compartment. A trial was instigated when the experimenter altered the illumination differential in such a manner as to make the stimulus compartment clearly visible from the conditioning compartment. The shock-unconditioned stimulus was automatically controlled by an Agastat relay. The response bar, mounted on the one-way vision screen, was modified slightly from that reported previously (7, 8). The essential modification was that the bar for the present experiment required a downward thrust instead of a pulling response. This modification was made primarily for additional studies and was only incidental to the present experiment. Dominance tests were conducted in the Wisconsin General Test Apparatus (WGTA) described by Harlow and Bromer (2).

Procedure. The *Ss* were hand-tamed daily in the WGTA for eight weeks prior to the beginning of the investigation. The daily taming trials were arbitrarily terminated after 25 raisins had been taken by the animal or after 15 minutes. Hand-taming was continued until such time as all *Ss* were responding satisfactorily in the test situation. The animals were all housed separately and were contained in closed-side cages to eliminate interactions with other members of the group outside of the experimental situation.

Following adaptation training the dominance structure of the entire group was determined. The dominance relationships were determined by placing a pair of animals in the restraining cage of the WGTA. The test was instigated by prefeeding each monkey with a raisin. On the succeeding ten trials a single raisin was placed on the moving tray of the WGTA and the attention of the animals drawn to it by tapping. The tray was then moved within reach of the animals and the experimenter recorded which animal obtained the food. The animal obtaining the greater number of raisins was designated as the dominant animal. None of the moving screens on the WGTA was used throughout the experiment, i.e., the experimenter was in full view throughout the test. The dominance testing procedure has been more fully described elsewhere (4). Each subject was paired with all other monkeys in the group. Tests were conducted at the rate of 10 pairs a day for four days with five pairings on the fifth day. The dominance structure was retested for the entire group starting two weeks after the beginning of the first test. Two weeks after the second group-dominance test was started, the animals were ranked on the basis of the two group tests. Adjacent animals were then tested in the dominance situation for five consecutive days. Thus, on each of the five pairs there were a total of seven dominance tests. The monkey which obtained the larger number of raisins on the greater number of the seven tests was designated as the dominant member of the pair.

In the interanimal conditioning phase the submissive animal was the conditoned stimulus for conditioning an avoidance response in the dominant animal. During conditioning the dominant member of each pair was placed in the conditioning compartment while the submissive member of the pair was placed in the stimulus compartment. On a conditioning trial the experimenter switched the lighting arrangement so that the stimulus animal was clearly visible from the conditioning compartment. Five seconds following the exposure of the stimulus animal, a 50-volt electric shock was automatically delivered through the grid floor to the animal being conditioned. Pressing the response bar enabled the animal to escape the shock or to avoid it if the response occurred within a five-second interval. The instrumental response also removed the stimulus monkey from view, i.e., response-terminated stimuli were employed. Twenty trials a day were administered with an average of 60 seconds between trials. Ten conditioning and 10 control trials were presented in a randomly interspersed pattern each day. On control trials the conditions were identical with the exceptions that the stimulus animal was removed from the apparatus and the trials were terminated manually after five seconds. No shock was present on the control trials. Conditioning was continued to a criterion of 90 per cent conditioned responses for three successive days with not more than 10 per cent responding on control trials for the same three days.

Twenty-four hours after the conditioning subject in each pair attained the conditioning criterion, dominance retesting in the WGTA was begun. The dominance status of the pair of monkeys was determined on five successive days. The data from the dominance tests following completion of conditioning appeared to be confounded somewhat by relatively poor adaptation to the WGTA. This difficulty was apparently due to the prolonged absence from the dominance test apparatus while conditioning was under way. In order to eliminate this source of contamination the animals were reconditioned after a 40-day interval. The original procedure was followed with the additional control that for 10 days preceding reconditioning and throughout reconditioning each animal was hand-tamed daily in the WGTA. This taming procedure consisted of presenting 12 single raisins on the moving tray of the WGTA. Each animal was tamed alone during this period. After attaining the reconditioning criterion, the five dominance tests for each pair were repeated. Two weeks after the five pairs of animals completed the experimental regimen, the dominance structure for the entire group was determined, i.e., every animal was tested with every other animal. The group dominance structure was reassessed after an interval of one week. One animal was excluded from the study prior to this

group testing since it appeared to be in ill health. This animal died some three months later.

On the basis of the interim group test, three additional pairs of monkeys were selected for study. Again an attempt was made to work with pairs of animals which were as nearly adjacent as possible in the dominance hierarchy. One animal selected to be conditioned to the sight of another monkey had previously been conditioned with another animal as the stimulus in an earlier pairing. The remaining two conditioning *S*s had served as stimulus animals in the first five pairings.

These three pairs of animals were given identical treatment to that described for the original pairings. That is, the dominance of each pair was tested for five successive days, conditioning was completed to criterion, and five paired tests were undertaken. Conditioning was then repeated as previously described with the inclusion of hand-taming trials during reconditioning, etc.

RESULTS

The dominance data for the submissive animal in each pair on the seven pretests (two group tests and five paired tests) are summarized in Table 1 along with the number of tests

TABLE 1
PERCENTAGE OF TESTS DOMINATED BY THE ORIGINALLY SUBMISSIVE MEMBER OF EACH PAIR

Pair	1	2	3	4	5	6	7	8
Pretests	0	7	0	29	43	36	0	36
Posttests	90	0	60	100	60	100	80	60

dominated by the submissive member on the five dominance measures which followed the completion of reconditioning. Data from the five dominance tests following original conditioning will not be reported. These data show the same trends but are less definitive.

The significance of changes in dominance following completion of conditioning was determined by computing a t with the difference technique (1) between the mean number of tests dominated by the submissive animal before and after conditioning. This comparison resulted in a t of 4.17 which with seven degrees of freedom is significant at the .01 level.

The acquisition of the original avoidance response was found to be very similar to that described in a previous study on interanimal conditioning (8). The number of conditioning trials required to attain the criterion ranged from 100 to 290.

DISCUSSION

The present study was designed in an attempt to verify and extend the observation that dominance relationships may be experimentally manipulated with the interanimal avoidance conditioning technique (5). After the dominant member of each of the eight pairs of animals had been conditioned with the submissive partner as the conditioned stimulus, the dominance relationship was reversed as predicted in seven of the eight pairings.

It is important to note that the conditioning tests and the dominance tests were conducted in separate and distinct situations. In addition to many distinctive features of the test situations, the experimental manipulations were conducted in separate rooms. In this case the learning in the conditioning situation generalized to a new situation and modified the subsequent behavior of the conditioned animals. This observation indicates that conditioning principles are not necessarily restricted to the artificial nature of soundproofed rooms and "pure" stimuli.

Not only did the CS acquire a new significance for the conditioned animal as is true of conditioning studies generally, but the CS in this instance acquired a meaning in an opposite direction to that existing prior to conditioning, i.e., a monkey which previously elicited dominant behavior now elicited submissive behavior.

This dramatic alteration in dominance behavior is somewhat surprising in that with the interanimal conditioning technique, the experimenter is only manipulating one aspect of a diadic relationship, i.e., the dominant animal presumably learns to become fearful at the sight of the submissive monkey. One might predict an enhanced effect if the submissive animal were taught in some manner to be nonfearful of the previously dominant animal. On several occasions it appeared clear that the previously dominant partner was not prepared to do combat to obtain the piece of preferred food; however the previously submissive animal in such pairings was also hesitant to take the food-getting initiative.

Although dominance determinations on the entire group of monkeys were important in determining the relative status of an animal in the group, it appeared desirable to work with single pairs of animals when the independent variable was administered. This modification in procedure from a previous study (5) made it possible to continue con-

ditioning to the same criterion for all animals. In addition, the interactions between dominance relationships which occur during group testing were eliminated by administering the critical dominance tests to pairs of animals.

The observed modifications in interindividual transactions conform very nicely with that suggested by the interpretation of avoidance conditioning which emphasizes the presence of fear as an intervening variable. The modification of dominance provides behavioral evidence for the presence of fear in avoidance learning which is independent of the conditioning situation. Recently completed experiments on interanimal conditioning in which the extinguished response in the previously conditioned animal is reinstated at a high level by observing the stimulus animal undergoing painful shock stimulation will make it possible to establish more definitive evidence for the presence of fear in avoidance conditioning and its communication between animals.

In the dominance tests following completion of interanimal conditioning the animals were carefully observed to note any deviation from the typical behavior of monkeys in this test situation. No differences could be distinguished, i.e., there was no apparent "bar-searching" activity, etc.

Since the modification in dominance following interanimal conditioning was attributed to the originally dominant animal learning to become fearful at the sight of the previously submissive animal, it was a logical extension to predict that as the conditioning was extinguished the dominance behavior would again revert to the original relationship. Due to the experimental design, only the last three pairs of animals could be used in determining this relationship; consequently, the data are more suggestive than definitive. In this phase of the experiment the extinction trials, following an additional reconditioning, were run each morning (except Sunday) and the dominance determinations were taken each afternoon.

The dominance pattern in all three pairs was found to be reversed to the original relationship as extinction training was continued. The number of conditioned responses during extinction and the number of raisins in the dominance test for a typical animal previously made dominant during the conditioning phase are presented in Fig. 1. Due to the startling correspondence between day-to-day fluctuations in the two experimental situations, the precaution was instigated of having one experimenter run the conditioning trials while the other conducted the dominance tests without any information about the conditioning data for that day; however, the basic relationship continued as before.

The relationship shown in Fig. 1 has interesting implications for the theoretical inter-

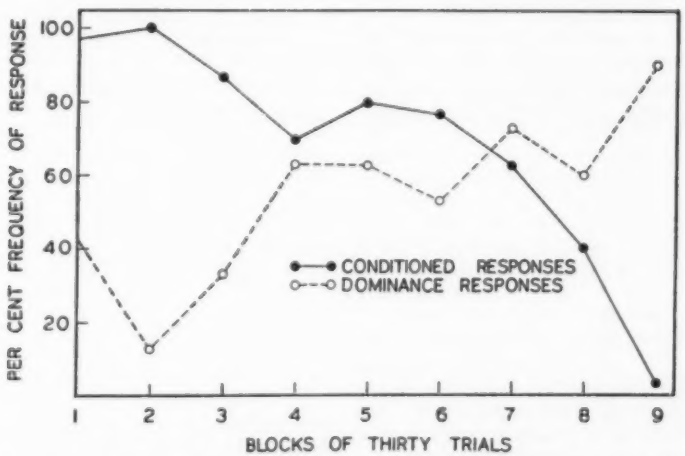

FIG. 1. CONDITIONED AVOIDANCE RESPONSES THROUGHOUT EXTINCTION AND RAISINS TAKEN IN THE DOMINANCE TEST BY A TYPICAL MONKEY

pretation of the extinction process of an avoidance response. In typical avoidance studies it is unknown whether the decrease in responding during extinction is attributable to extinction of the motor response, the fear response, or a combination of the two. The postulation (9) that the fear attendant to CS presentation decreases on the nonreinforced presentations during extinction with the CR subsequently diminishing in frequency is supported by the relationship depicted in Fig. 1. In this case an independent measure of the capacity of the CS to elicit fear at varying stages throughout extinction was available. Since the measure of fear in the social situation (number of raisins taken in the dominance test) increased at approximately the same rate that conditioned avoidance responses decreased, it suggests the inference that extinction of the bar-pulling response is primarily attributable to decreased motivation (fear) and not to fatigue accompanying the conditioned response, as many theorists would lead us to believe.

SUMMARY

Two group dominance tests were conducted on ten rhesus monkeys. On the basis of these tests five pairs of animals adjacent, or nearly adjacent, in the hierarchy were given an additional five dominance determinations. The animal in each pair which received the greater number of raisins in each of the seven tests between the two animals was designated as dominant. This animal in each pair was subjected to avoidance conditioning with his submissive partner as the conditioned stimulus. After attaining a conditioning criterion, each pair of monkeys was given an additional five dominance tests.

After completing this phase two additional group-dominance tests were completed. An additional three pairs were selected as before and subjected to the same procedure. The dominance status was found to be significantly reversed following the completion of conditioning. It was suggested that this observation provides behavioral evidence for the presence of fear in avoidance conditioning which is independent of the conditioning situation.

REFERENCES

1. EDWARDS, A. L. *Statistical analysis*. New York: Rinehart, 1946.
2. HARLOW, H. F., & BROMER, J. A. A test apparatus for monkeys. *Psychol. Rec.*, 1938, **2**, 434–436.
3. MASLOW, A. H. Dominance-quality and social behavior in infra-human primates. *J. soc. Psychol.*, 1940, **11**, 313–324.
4. MILLER, R. E., & MURPHY, J. V. Social interactions of rhesus monkeys: I. An examination of the stability of dominance relationships. *J. soc. Psychol.*, in press.
5. MILLER, R. E., MURPHY, J. V., & MIRSKY, I. A. The modification of social dominance in a group of monkeys by interanimal conditioning. *J. comp. physiol. Psychol.*, 1955, **48**, 392–396.
6. MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
7. MURPHY, J. V., & MILLER, R. E. A flexible general test apparatus for the monkey. *J. gen. Psychol.*, in press.
8. MURPHY, J. V., MILLER, R. E., & MIRSKY, I. A. Interanimal conditioning in the monkey. *J. comp. physiol. Psychol.*, 1955, **48**, 211–214.
9. MURPHY, J. V., MILLER, R. E., & FINOCCHIO, D. V. Spontaneous recovery of an avoidance response over an extended time interval in the monkey. *J. genet. Psychol.*, in press.

Received May 11, 1956. Prior publication.

CRITIQUE AND NOTES

AN INVESTIGATION OF PHONETIC SYMBOLISM[1]

IRVING MALTZMAN, LLOYD MORRISETT, JR.,[2] AND LLOYD O. BROOKS

University of California, Los Angeles

DURING the course of several years we have conducted a series of studies that are directly related to the experiments on "phonetic symbolism in natural languages" recently reported by Brown, Black, and Horowitz (1). They found, in accord with the work of earlier investigators, that *S*s could match foreign words with their English equivalents significantly beyond chance expectancy. The languages employed were Chinese, Czech, and Hindi. They interpreted their results in terms of a universal phonetic symbolism common to natural languages. This universal phonetic symbolism is based upon the hypothesis of physiognomic language which assumes that there are unlearned interconnected sensory systems underlying sounds and meanings, an application of the gestalt principle of organization that communication may occur among different trace systems that are organized in a similar fashion. Foreign words can be matched with their English equivalents because they are both related to similarly organized neural trace systems.

The principle of gestalt organization was invoked by Tsuru and Fries (2) to account for their finding that *S*s could select the Japanese equivalents of English words with a high degree of success. The results of the present experiments confirm and extend the empirical results of these previous investigators by obtaining similar results with a foreign language not previously employed. However, it is believed that certain of our results cast doubt upon the foregoing interpretation of the phenomena.

The present experiments were all conducted in group form. The *S*s were students in the senior author's psychology classes. At the beginning of a class hour, *E* would distribute questionnaires to the students and make a brief announcement requesting their cooperation in the investigation. The questionnaire contained 25 stimulus words and 25 pairs of response words, one pair for each stimulus. A sample of the instructions are as follows:

> Below is a list of Japanese and English words. For each Japanese word I want you to select the corresponding English word. Of course you do not know the meaning of the Japanese words. But try to pick the English word that seems to best fit the Japanese word.
>
> In order for you to be successful, it is necessary that you avoid artificial terminological schemes such as associating sounds, letters, length of words, etc.
>
> Underline the word that you think is correct. *Please do not ask any questions or call out until the papers are collected.*

No further instructions or illustrations concerning the pronunciation of the foreign words were given.

The *S*s were allowed to work on their answers for as long as they wished. The time required was not more than 10 minutes. The *S*s were also asked to indicate whether they had any acquaintance with the foreign language(s) in question. The results of seven *S*s were discarded from the various experiments. Six *S*s were of Japanese descent and one had spent several years in Japan in the Army of Occupation.

The first experiment was designed to determine whether the order of the English and the foreign words, Japanese in this case, and the nature of the student sample would influence the frequency of choice of the equivalents. In the Japanese-English mode of presentation, the 25 Japanese words were the stimuli and English words were the response pairs. One member of each pair was the equivalent of the Japanese stimulus word. In the English-Japanese mode of presentation, 25 English words were the stimuli and Japanese words formed the response pairs. The first and second columns in Table 1 contain the English words and the Japanese equivalents employed in this and subsequent experiments. All of these words were obtained from the article by Tsuru and Fries (2). The Japanese stimulus words for the first list were randomly selected from each pair of antonyms in their list. The order of the correct and incorrect member of each response pair was varied randomly from pair to pair. The English stimulus words in the second list were the equivalents of the Japanese words used in the first list.

A total of 65 *S*s served in this experiment. There were 25 undergraduate and 40 graduate *S*s in the Japanese-English condition, and 16 undergraduate and 19 graduate *S*s in the English-Japanese condition. Mood's nonparametric median test of the number of errors committed by the four different groups was not significant, $\chi^2 = 3.29$, $df = 3$, $p > .30$. The groups were therefore combined.

[1] This study was aided by a grant from the Committee on Research, Graduate Council, University of California.

[2] Now at Yale University.

TABLE 1

ENGLISH WORDS AND THEIR JAPANESE AND CROATIAN EQUIVALENTS

	English	*Japanese*	*Croatian*		*English*	*Japanese*	*Croatian*
1.	bird	tori	ptica	13.	high	takai	visok
	worm	mushi	crv		low	hikui	nizak
2.	red	aka	crven	14.	kite	take	zmaj
	green	midori	zelen		boat	hune	čamac
3.	peace	heiwa	mir	15.	old	oitaru	star
	war	tatakai	rat		young	wakai	mlad
4.	sweet	amai	sladak	16.	hot	atsui	topal
	bitter	karai	gorak		cold	samui	hladan
5.	fast	hayai	brz	17.	are	aru	jesu
	slow	osoi	spor		are not	nai	nisu
6.	white	shire	bijel	18.	blue	ao	plav
	black	kuro	crn		yellow	kii	žut
7.	square	shikaku	četvorina	19.	thick	atsui	debel
	circle	maru	krug		thin	usui	tanak
8.	good	yoi	dobar	20.	big	ookii	velik
	evil	warui	zao		small	chiisai	malen
9.	praise	homeru	hvaliti	21.	clear	sunda	jasan
	deprecate	kenasu	kuditi		muddy	nigotta	mutan
10.	far	tooi	dalek	22.	enemy	mikata	dušman
	near	chikai	bliz		friend	teki	prijatelj
11.	soft	yawai	mekan	23.	crooked	magatta	savijen
	hard	katai	tvrd		straight	massugu	ravan
12.	smart	riko	pametan	24.	right	tadashii	prav
	dull	baka	glup		wrong	ayamereru	kriv
				25.	sharp	togatta	oštar
					dull	nibui	tup

The mean number of errors obtained by these *S*s was 11.11, $\sigma = 2.67$, where 12.50 errors would be expected by chance. The obtained divergence from chance is highly significant, $CR = 4.21$, $p = < .0001$.

The results of this experiment confirm the findings of Tsuru and Fries (2) and indicate that the order of presentation of the English and Japanese equivalents does not have a significant effect. Academic level, likewise, did not influence significantly the frequency of choice of the correct equivalent.

As Brown *et al.* (1) point out, however, the lists employed in this first experiment were constructed by Tsuru who was familiar with both Japanese and English, thereby providing a possible source of bias. This difficulty is avoided in the next experiment where Croatian equivalents of the English stimulus words of the English-Japanese list were employed.[3] Forty-eight *S*s received a list of 25 Croatian stimulus words with adjacent pairs of English response words, one of which was the equivalent. The Croatian words used in this and in subsequent studies are given in the third column of Table 1.

The obtained mean number of errors in this experiment was 11.31, $\sigma = 1.97$, $CR = 4.25$, $p = < .0001$. These results clearly indicate that the English equivalents of Croatian words can be selected with a frequency significantly beyond chance expectancy, and quite as effectively as the selection of English and Japanese equivalents.

The results of these two experiments are in accord with the empirical findings of Brown, *et al.*, (1), and Tsuru and Fries (2), and appear to support their hypotheses of phonetic symbolism and gestalt organization of trace systems.

If the above theoretical interpretation is correct, however, the presence of an English word in a stimulus-response pair is not essential for a correct match. The *S*s given pairs of Japanese and Croatian words should be able to select the appropriate equivalents as readily as in the case with English and foreign words, since the interconnections or communication pathways among the sensory systems related to the equivalents are unlearned and universal.

In the third experiment, therefore, Croatian words were used as the stimulus words with Japanese response words, one member of each pair being an equivalent of the Croatian word. Twenty-nine *S*s served in this experiment. The mean number of errors obtained was 12.28, $\sigma = 2.57$, $CR = .46$, $p = > .60$.

In order to make certain that the order of pres-

[3] We are indebted to Mr. Lotar Zahradka for providing the Croatian equivalents.

entation was not a variable responsible for the lack of significance in this experiment, a fourth experiment was conducted. Forty *Ss* were given Japanese stimulus words with Croatian response words. The results were the same as in the previous experiment. A mean of 12.15 errors was obtained, $\sigma = 2.76$, $CR = .80$, $p = > .40$.

The results of the last two experiments therefore appear to be contrary to the gestalt principle of organization and the notion of a universal phonetic symbolism. It seems that for matching to be successful by English-speaking subjects, one member of the match must be an English word. This implies that the ability to match foreign equivalents is based upon complex kinds of learning, involving, we would surmise, mediated generalization.

SUMMARY AND CONCLUSIONS

Four experiments were conducted investigating the accuracy with which *Ss* can correctly match English words and foreign equivalents. The first experiment conducted showed that *Ss* could match English and Japanese equivalents at a level beyond chance expectancy, and that there was no significant difference between undergraduate and graduate *Ss* in this respect. A second experiment showed that English and Croatian words could be matched at a level beyond chance expectancy. However, two additional experiments showed that the correct equivalents among Japanese and Croatian word pairs were not matched at greater than a chance level. The latter finding was interpreted as being contrary to the hypothesis of the gestalt organization of trace systems, and the related hypotheses of phonetic symbolism and physiognomic language.

REFERENCES

1. BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages. *J. abnorm soc. Psychol.*, 1955, **50**, 388–393.
2. TSURU, S., & FRIES, H. S. A problem in meaning. *J. gen. Psychol.*, 1933, **8**, 281–284.

Received August 29, 1955.

PSYCHOTHERAPY AND THE RECOVERY FROM NEUROSIS

DALBIR BINDRA

McGill University

BOTH practically and theoretically, it is important to find out whether or to what extent psychotherapy is effective in the treatment of various behavioral disorders. This paper considers from a critical and broad viewpoint the problem of the efficacy of psychotherapy in the treatment of neurosis. It separates and clarifies the specific issues involved, and attempts to show what questions can or cannot be answered on the basis of available evidence.

PROBLEM

Basically, the question is whether or not the proportion of recoveries is greater among neurotics who undergo psychotherapy than among neurotics who do not receive such therapy. Meehl, in his recent review of the literature on this subject (5), notes the lack of the type of controlled studies which would unequivocally answer this question. In the absence of adequate experimental evidence on this question, psychologists have tried to argue (e.g., 2, 3, 8) for or against the efficacy of psychotherapy on the basis of indirect evidence obtained from improperly controlled studies. In general these arguments reach one of two conclusions: "No one has yet demonstrated that psychotherapy is effective in treating neurosis," or "No one has yet demonstrated that psychotherapy is *not* effective in treating neurosis." Clearly, this question is amenable to experimental attack. With two groups of patients equated in terms of symptoms, severity of neurosis, motivation for therapy, socioeconomic status, intelligence, and the like, only one of the groups must receive psychotherapy. A comparison of the proportion of recoveries in the two groups will then indicate whether psychotherapy is effective.

Unfortunately, discussions of this rather straightforward empirical question can easily get complicated by certain theoretical preconceptions. In the thinking of many psychologists, the problem of the efficacy of psychotherapy is intimately tied up with the problem of the etiology of neurosis. The notion that life experiences play a dominant and crucial role in the onset of neurotic ailments is generally considered to support the view that psychotherapy can cure these ailments. And the view that neuroses arise primarily from organic or nonpsychological factors is considered to favor the view that psychotherapy is ineffective. No one has explicitly stated and defended this type of argument, but it is implicit in most reputable psychological discussions (e.g., 9). There is little justification, however, for this contention

that etiology and therapy bear such a direct and clear-cut relation to each other. The only effective treatment of a disorder of organic origin (like cerebral palsy) may consist of some form of training or psychotherapy; on the other hand, a psychogenic disorder (like some insomnias) may be most effectively treated by medication. The mere knowledge of the etiology in these cases does not tell whether a particular treatment would be effective. Similarly, whether neurotic ailments result from psychological conflicts, environmental stresses, or some subtle chemical factors in the blood is quite irrelevant in determining the efficacy of psychotherapy in their treatment. I do not mean to imply that neurotic disorders can be identified in a way that excludes etiological concerns. Rather, the point here is simply that the effectiveness of psychotherapy is an issue that can be and should be decided without linking it with any specific conjecture about the etiology of neurotic ailments.

Clarity and explicitness are also necessary in defining neurosis. In the present discussion, *neurosis* or neurotic disorder refers to cases of persistent and gross maladjustments involving some definite, palpable, behavioral peculiarities (or symptoms) such as anorexia, phobias, compulsive acts, demonstrable anxiety attacks, hysterical blindness, impotence, amnesias, and the like. It should be noted that this rough and arbitrary definition excludes minor maladjustments defined with reference only to subjective states such as "unhappiness," "anxiety," and "tension." It is true that "anxiety," "unhappiness," and "tension" are often (though not always) reported by neurotics, but they are also frequently reported by normals. Since these states are not a *sine qua non* of persistent and gross maladjustments unless accompanied by definite symptoms of the type listed above, they are excluded from the present definition of neurosis. Some psychologists believe that both neurosis and these minor maladjustments arise from the same type of etiological factors. Even if this view is correct, it does not necessarily imply that an effective treatment of minor maladjustments would also be effective in cases of neurosis, or vice versa. The concern here is primarily with recovery from neurosis.

A General Formulation

Recovery from neurosis means nothing more than a kind of behavioral change, a change in those aspects of the patient's behavior that put him in the category of neurotics. Thus, the question of the effectiveness of psychotherapy is really a special case of the general problem of the extent to which psychotherapy can produce *personality change*. A clear formulation of this question requires a closer analysis of the concepts of "personality change" and "psychotherapy." Precise descriptions are needed of (*a*) the personality variables along which change is presumed to occur, and (*b*) the psychotherapeutic processes which are presumed to effect the change. The effect of a specified therapeutic process on a specified personality variable can then be investigated.

Personality variables can be specified easily enough. For present purposes, they can arbitrarily be categorized into five rough categories: (*a*) *cognitive variables*, such as general information, reasoning ability, and memory; (*b*) *attitude variables*, such as attitudes toward in- and outgroups, attitude toward oneself, and likes and dislikes; (*c*) *need variables*, such as introversion, super-ego, anxiety, and aggression; (*d*) *specific neurotic symptoms*, behavioral characteristics, such as hysterical or compulsive symptoms, which are used in defining persons as neurotics; (*e*) *personality variables (if any) that are causally related to neurotic ailments*. Intelligence tests, scales of attitudes, and "personality tests" could be used to measure changes along the first three types of personality variables. The last two categories are not at the same level of discourse as the first three, nor have they been adequately identified so far, but they are listed here to focus attention on the type of variables that are of direct concern in this discussion. Accepting this admittedly arbitrary categorization, the problem of the effect of psychotherapy on personality variables resolves itself into five separate questions. Does psychotherapy have any effect on (*a*) cognitive variables, (*b*) attitude variables, (*c*) need variables, (*d*) specific neurotic symptoms, and (*e*) personality variables (if any) that cause neurotic symptoms?

To answer these questions, it is necessary to specify the exact nature of psychotherapy.[1] Unfortunately, it is not possible to do so. Psychotherapy still has its mysterious aspects, and present knowledge is far from unraveling the essential nature of all that takes place in the therapist's office. But from all accounts, it appears that the therapist provides the patient with a friendly, permissive, and uncritical atmosphere, and then systematically makes use of "interpretation," suggestion, and catharsis. It is not clearly understood, however, exactly how these various processes operate in the psychotherapeutic situation. Ideally, empirical questions should be formulated in terms

[1] This conception excludes the *incidental* "psychotherapy" that is sometimes said to have taken place when a neurotic gets some attention from a nurse or when he confides his innermost thoughts to his bartender, priest, or barber. Whether or not incidental "psychotherapy" has some of the same ingredients as psychotherapy proper (given by a qualified psychiatrist or clinical psychologist) is impossible to say until more is known about what is involved in the latter.

of specific psychotherapeutic processes. Thus, one could ask such questions as "Can suggestion change attitudes and needs?" or "Can interpretive analysis affect compulsive or hysterical symptoms?" or "Can catharsis relieve functional impotence?" Such questions, however, cannot be answered unless psychotherapy is analyzed into its component processes rather than conceived as a unitary whole. The only point here is that if knowledge of exactly how psychotherapy produces whatever effects it does produce is to be gained, then an analysis of this type must be undertaken. From the point of view of the therapist, psychotherapy may well be considered as a unitary healing device, but for the researcher it is an analyzable complex of psychological processes. For the present, however, one is forced to discuss the issues without benefit of this kind of analysis. With this handicap in mind, questions concerning the effects of psychotherapy on the various categories of personality variables may now be considered.

EFFECTS OF PSYCHOTHERAPY

Inasmuch as psychotherapy involves prolonged social interaction between the therapist and the patient, it is to be expected that the psychotherapeutic situation will produce change along at least some personality dimensions. It is known that interpersonal relations can affect behavior. Interests and tastes change through association with friends. Gestures of immigrants begin to conform to the typical gestures of their adopted communities (1). Exposure to conflicting social influences may contribute to behavioral disturbances in juveniles (4). Attitudes can be changed, one way or the other, simply by showing a short movie to audiences (6). In view of the abundance of naturalistic and experimental evidence of this kind, it would be strange, indeed unbelievable, if the long and intimate social interaction between the therapist and his patient did not produce some change in the patient (and, indeed, in the therapist himself). The patient is likely to learn something about the work of the therapist, to increase his vocabulary, and to change his attitudes. Since he is the focus of attention in the therapeutic situation, the patient is particularly likely to learn more about himself and to evaluate himself rather differently. Also, the change from a nagging wife, a rude boss, and critical friends to the understanding and permissive therapist is likely to facilitate these new self-evaluations, and they may be more in keeping with the therapist's direct or indirect suggestions regarding what a normal person thinks of himself.

The recent studies of Rogers, Dymond, and their collaborators (7) present direct evidence in support of these statements. They administered a number of personality tests to their patients before, during, and after psychotherapy. Comparable control data were obtained from normal subjects, who were not given therapy, and from patients who had to wait for therapy. The most impressive finding was that psychotherapy produced a significant change in self-perception or in the perception of the self-ideal. It was also found that responses of patients to the Thematic Apperception Test changed as a result of psychotherapy in the direction of greater judged personality integration and adjustment. The results on needs and changes in behavior in everyday life situations were equivocal, and no change in attitudes toward others could be attributed to the treatment. No data are presented on the changes in specific neurotic symptoms. It seems appropriate to conclude that psychotherapy has some effect on the client's attitudes toward himself. That specific psychotherapy (as well as informal interpersonal relations generally) can produce behavioral changes does not, of course, diminish the need for analyzing "psychotherapy" into its component processes, as suggested in the last section. If one could identify the specific effective agents in the psychotherapeutic situation, it would be a step toward more efficient therapeutic procedures.

These demonstrations by Rogers, Dymond, and collaborators, important as they are, do not, of course, tell anything about the efficacy of psychotherapy in the cure of neurosis. These workers have shown that the attitudes of their patients changed, and that, *on the basis of test data*, they were said to become more "mature" and "better adjusted." But this does not indicate whether psychotherapy was effective in eradicating the specific neurotic symptoms of their clients. Even if psychotherapy demonstrably alleviates some minor maladjustments involving attitudes of different kinds, the issue of the effectiveness of psychotherapy in the treatment of the basic, gross, and persistent maladjustment ("primary neurotic symptoms") still remains unanswered. A neurotic with hysterical or compulsive symptoms may well become better adjusted as a result of psychotherapy without showing any change in his primary neurotic symptoms. Even a person who loses a leg may benefit from psychotherapy inasmuch as his life-goals and attitudes may change in a way that would be more consistent with his handicap, and in this sense he may become more mature, better adjusted, and happier. However, in order to demonstrate that psychotherapy is effective in curing neurosis, it must show that it diminishes or eradicates the symptoms that constitute the neurosis, quite apart from making the patient better adjusted, mature,

or happier. Such a demonstration has not yet been made. Perhaps future research will show that certain psychotherapeutic processes can be effective in the treatment of certain types of neurotic maladjustments.[2]

Even if it were demonstrated that psychotherapy can relieve neurotic ailments, there would still be no answer to the final question. Can psychotherapy produce a change in those processes or personality variables (if any) that are causally connected with neurosis? Evidence showing psychotherapy to be effective in eradicating neurotic symptoms would not indicate anything about the mechanism by which psychotherapy operated. In this connection, two types of treatments should be distinguished. One kind involves directly undoing the disease process analogous to treating streptococcal infections with penicillin. The other type of treatment is indirect and is aimed at something other than the disease process itself; an analogy is treating typhoid or tuberculosis with rest. In textbooks on psychiatry and abnormal psychology, it is generally implied that the effectiveness of psychotherapy results from directly undoing (through relearning) the processes that produce neurosis. However, there is no evidence for this belief, and there cannot be until we know the exact basis of neurosis. It is quite conceivable that, were psychotherapy shown to be effective in treating neurosis, its effectiveness could result from some indirect effects that are quite unrelated to the causes of neurosis.

In summary, available evidence suggests that psychotherapy can be effective in alleviating minor maladjustments characterized by such subjective states as "unhappiness," "anxiety," and "tension" that often accompany neurosis. But the evidence does not yet support the view that psychotherapy is effective in relieving those gross and persistent maladjustments (neuroses) that are characterized by definite, palpable symptoms. Whether it has any effect on these is a question which only future research can settle.

[2] Whether the recovery from neurosis could be more effectively produced by methods other than psychotherapy is an important question, but one that need not be raised in the present discussion.

REFERENCES

1. EFRON, D., & FOLEY, J. P. Gestural behavior and social setting. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 33–40.
2. EYSENCK, H. J. The effects of psychotherapy: an evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.
3. EYSENCK, H. J. The effects of psychotherapy: a reply. *J. abnorm. soc. Psychol.*, 1955, **50**, 147–148.
4. HUNT, J. McV. An instance of the social origin of conflicts resulting in psychoses. In C. Kluckhohn & H. A. Murray (Eds.), *Personality in nature, society and culture*. New York: Knopf, 1948. Pp. 367–374.
5. MEEHL, P. E. Psychotherapy. *Annu. Rev. Psychol.*, 1955, **6**, 357–378.
6. PETERSON, R. C., & THURSTONE, L. L. *Motion pictures and the social attitudes of children*. New York: Macmillan, 1933.
7. ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
8. ROSENZWEIG, S. A transvaluation of psychotherapy—a reply to Hans Eysenck. *J. abnorm. soc. Psychol.*, 1954, **49**, 298–304.
9. WHITE, R. W. *The abnormal personality*. New York: Ronald, 1948.

Received September 3, 1955.

A SHORT FORM OF WITKIN'S EMBEDDED-FIGURES TEST[1]

DOUGLAS N. JACKSON
The Menninger Foundation

SINCE Gottschaldt (2) studied the role of past experience in perception by providing differential experience with simple forms and later determining how well they were masked when contained in a more complex pattern, various modifications (e.g., 3, 4) of the original patterns have been developed to measure individual differences in perceiving. Recently Witkin (5) has developed an embedded-figures test, based on the Gottschaldt figures, but more difficult than the latter due to the addition of color to the complex patterns. In an extensive study of perception-personality relationships by Witkin and his colleagues (6) this test proved to correlate significantly with measures of orientation to the upright, as well as with a host of personality variables. Sex differences in scores on this test, favoring males, have consistently emerged.

Of the perceptual variables used in the Witkin

[1] This research was completed while the author was a Public Health Service Postdoctorate Research Fellow at The Menninger Foundation.

studies, the embedded-figures test is the only one which does not require expensive apparatus, and hence will probably be the one most frequently used by other research workers. A major difficulty exists in administering this test. Since the score is the time required to extract simple figures from each of 24 complex patterns, some *S*s require an excessively and unpredictably long time to complete the test—sometimes more than an hour and a half.

An item analysis was conducted with the aim of reducing the length of time required to administer the test without substantially altering what the test measures. The embedded-figures test was administered to 50 college students, 19 men and 31 women, as a part of a larger study (1). From this group, 26 *S*s, 13 from each extreme of the distribution, were selected. Twenty-four single-classification analyses of variance, one for each pattern, were conducted. The 12 items which most significantly (all at better than the .01 level) discriminated between the extreme groups (and hence showed the highest relationship with the whole scale) were selected for inclusion in a short form. The following patterns comprise the short form, using Witkin's (5) numbering system: C-1, D-1, E-1, A-2, C-2, G-1, A-3, H-1, E-3, C-3, D-2, and E-5.

The short form correlates .99 with the whole scale when the original 50 *S*s are used. In a cross validation with Witkin's (5) original group,[2] the short-form—full-scale correlations are as follows: men (N = 51), 96; women (N = 51), .96. In an additional cross validation with a group of *S*s given the embedded-figures test, as part of an as yet unpublished study conducted at The Menninger Foundation, the short-form—full-scale correlations are: men (N = 30), .98; women (N = 30), .98.

[2] The author is grateful to Dr. H. A. Witkin for making these data available.

Although reducing the number of items from 24 to 12 effected more than a 40 per cent reduction in administration time, the possibility of a further reduction remained. Since eliminating additional items could reduce the reliability, a reduction in the time limit for each item from five to three minutes was considered. In the Menninger group the correlations between the short form with the three-minute time limit and the original scale are: men, .96; women, .97. The three-minute time limit may have some advantage in that the distribution of items has a less extended tail and is not as asymmetric as the distribution of each item with the five-minute time limit.

It thus appears that the 12-item short form with a three-minute limit gives a very good approximation of results obtained from the entire Witkin embedded-figures test while requiring only half as much administration time.

REFERENCES

1. JACKSON, D. N. Stability in resistance to field forces. Unpublished doctor's dissertation, Purdue Univer., 1955.
2. GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren, I: Über den Einfluss gehäufter Einprägung von Figuren auf ihre Sichtbarkeit in unfassenden Konfigurationen. *Psychol. Forsch.*, 1926, **8**, 261–317.
3. GUILFORD, J. P., & LACEY, J. I. (Eds.) *Printed classification tests.* Washington: U. S. Government Printing Office, 1947.
4. THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. of Chicago Press, 1944.
5. WITKIN, H. A. Individual differences in ease of perception of embedded figures. *J. Pers.*, 1950, **19**, 1–15.
6. WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

Received September 26, 1955.

ACHIEVEMENT MOTIVATION AND LEARNING

PATRICIA A. KAROLCHUCK AND LEONARD WORELL
Portland, Oregon *Reed College*

THE purpose of this study was to investigate the effects of n Achievement on two kinds of learning, directed and incidental. A successful attempt was recently made by Lowell (1) to predict degree of learning from the McClelland n Achievement index. He found that a group with high n Achievement increased its output more from beginning to end of a Scrambled Words Task than a group with low n Achievement. Our intent here was to re-examine Lowell's findings using altered learning materials and to extend our investigation to the effects of n Achievement on incidental learning. From the hypothesis that increased achievement motivation leads to increased learning, we anticipated that a high n Achievement group, as compared with a low n Achievement group, would manifest a greater degree of both directed and incidental learning.

METHOD

Materials. The learning task consisted of an adaptation of a James Thurber short story which was unfamiliar to the subjects (*Ss*). A multiple-choice test was developed to determine the degree of learning of the story. A similar test was employed by Postman and Senders (3) who investigated the influence of set in a learning situation by providing instructions which centered attention upon either general or specific aspects of learning material. They assumed that the effects of an explicit set were tested when *S* was asked a question that he had been prepared to expect. Further, when *S* was questioned on matters unrelated to this set, the effects of incidental learning were being examined. Our multiple-choice test comprised 20 questions, ten each dealing with details of content and general comprehension.

Subjects and procedure. The McClelland group TAT, comprising four pictures,[1] was administered to 108 high school freshmen under neutral conditions. Each picture was projected on a screen for 20 seconds, after which *S* was allowed four minutes to write a story.

The TAT stories were scored by two judges (interrater $r = .96$). The 12 highest and 14 lowest *Ss* in n Achievement were used to test the hypotheses. These groups did not differ in age, IQ, grade average, or reading ability. Prior to reading the short story, *Ss* were told to look for spelling and grammatical errors and details of content. In order to insure a single reading of the story, each *S* recorded his total reading time and handed in his test immediately upon completion.

RESULTS AND DISCUSSION

Contrary to Lowell's findings (1), we have failed to support an anticipated relationship between n Achievement and directed learning. Individuals with High n Achievement did not perform more effectively than Low n Achievement scorers in a situation in which a particular kind of learning was the primary induced goal (Table 1). There was, however, substantial evidence for the presence of a relationship between n Achievement and learning that was irrelevant to the induced set. High n Achievement individuals manifested more learning of an incidental nature than those with Low n Achievement (Table 1).

The discrepancy between Lowell's experiment and the present one may be attributed to a difference in the nature of the learning situations. In the Scrambled Words Task utilized by Lowell, performance could be improved by practice and increased effort, while in this study no opportunity for improvement was provided. This explanation, however, does not seem adequate in the light of the occurrence of superior incidental learning by High n Achievement scorers in this research. For the present, we believe that our results must be accepted as empirical findings because of the unclear composition of n Achievement. In re-examining the referents for n Achievement, at least two distinct components appear to be involved, an experience and a motivational variable.[2] McClelland *et al.* (2) seem to suggest this possibility in referring to motives as comprising expectations, which are "... based on repeated exposure to certain stimuli ..." (2, p. 76), and achievement interest or desire. Though the terminology may not be congenial, the essential point remains that several variables may be operant in the n Achievement Scale, whose antecedent conditions are unclear and whose relationship to one another is unspecified. Consequently, it would seem necessary to examine the McClelland index with respect to the nature of the variables involved and to make provisions for their control in order to maximize prediction.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF HIGH AND LOW n ACHIEVEMENT GROUPS IN DIRECTED AND INCIDENTAL LEARNING CONDITIONS

Learning Condition	Low n Ach.		High n Ach.		t	p
	Mean	*SD*	Mean	*SD*		
Directed	5.21	9.73	5.75	8.02	.12	.50
Incidental	4.07	9.80	5.83	7.80	1.87	$< .05$

SUMMARY

This study was designed to determine the relationship between n Achievement and learning. It was hypothesized that individuals with high n Achievement would learn more in directed and incidental learning situations than those with low n Achievement.

The findings may be summarized as follows:

[1] For a more detailed description of the materials and procedure, see McClelland *et al.* (2). The following four pictures were used:

a. two men in a shop working at a machine (3, p. 100)

b. boy with an open book in front of him at a desk (3, p. 101)

c. an older man and a younger man (TAT 7 BM)

d. young boy standing before a surgical mural (TAT 8 BM)

[2] For example, two subcategories, obstacles or blocks (Bp, Bw) and achievement thema (Ach Th), are each given credit toward indicating n Achievement. It would seem to the authors that "obstacles or blocks" may be related to *expectancies* for nonattainment of motive satisfaction, whereas "achievement thema" reflects *concern* over goal achievement but not necessarily expectancy of occurrence.

(*a*) The expected superiority of High n Achievement *S*s in a directed learning situation was not found. Our results are not in agreement with Lowell's findings. (*b*) High n Achievement scorers on the other hand, did demonstrate more efficient incidential learning. The present difficulty in interpreting our results was considered to be a function of the unclear specification of the several possible variables constituting the n Achievement Index, and of their interrelationships.

REFERENCES

1. LOWELL, E. L. The effect of need for achievement on learning and performance. *J. Psychol.*, 1952, **33**, 31–40.
2. MCCLELLAND, D., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. The achievement motive. New York: Appleton-Century-Crofts, 1953.
3. POSTMAN, L., & SENDERS, VIRGINIA. Incidental learning and generality of set. *J. exp. Psychol.*, 1946, **36**, 153–165.

Received October 11, 1955.

THE CHILD'S PERCEPTION OF THE PARENT

JEROME KAGAN

The Ohio State University

MUCH of the contemporary research on parent-child relationships has centered on the child's perception of his parents. It is generally assumed that acquisition of behaviors like dominance, submission, and other response patterns involved in identification, is influenced by the child's perception of the role-behaviors of each parent figure. However, because of the lack of relevant empirical data on a normal sample it is difficult to (*a*) assess the significance of the perceptions of a pathological sample and (*b*) evaluate the relative merit of the various hypothetical statements relating parental behavior to personality development.

Funkenstein, King, and Drolette (1) administered to a group of Harvard students a questionnaire designed to reveal the subject's perception of his parents. The most frequent perception was that of the father as chief source of authority and the mother as major source of affection. It was emphasized that Harvard students are a select sample and caution must be exercised in generalizing these findings. The present brief report summarizes the answers of a much different population, a group of Ohio school children, to direct questions tapping attitudes toward both parents.

METHOD

Interviews were held with 217 children (106 girls and 111 boys), ages 6–0 to 10–2, from grades one, two and three in a Columbus, Ohio, public school. Their answers to the following questions were obtained.

1. If you were in an argument at home with your mother and father, who would be on your side, your mother or your father?
2. Let's make believe you were bad and your mother and father were both home. Who would punish you, your mother or your father?
3. Who is the boss in your house, your mother or your father?
4. Who are you scared of more, your mother or your father?

RESULTS AND DISCUSSION

Table 1 shows the percentage of boys and girls answering "Mother" to each of the four questions. The majority of boys and girls perceived mothers as friendlier, less punitive, less dominant, and less threatening than fathers ($p < .01$ for each question; all p values are for two tails). These results are in essential agreement with those of Funkenstein, King, and Drolette (1).

TABLE 1

PERCENTAGE OF BOYS AND GIRLS RESPONDING "MOTHER" TO EACH OF THE FOUR QUESTIONS

Group	*N*	Question 1	Question 2	Question 3	Question 4
Boys	111	60.9	36.4	27.9	16.7
Girls	106	73.6	33.0	19.8	29.2

Although one reasonable explanation of the data is that these answers reflected the child's perception of the objective family situation, it is possible that the statements of both the Harvard students and the Ohio children were influenced, in part, by a culturally reinforced attitude that fathers *ought to be* less affectionate and more dominant than mothers. This hypothesis is tentatively supported by the children's verbal elaborations to the questions. For example, to the question "Who is the boss in your house . . . ?" many *S*s hesitated before answering and if they said "Father" often explained their answer by adding "Because he has to work for a living" or "Because he owns the house."

TABLE 2

PERCENTAGE OF BOYS AND GIRLS RESPONDING "MOTHER" TO THE QUESTIONS IN RELATION TO GRADE PLACEMENT

Question	Grades 1 and 2		Grade 3	
	Girls	Boys	Girls	Boys
1	75.7	54.9	69.4	66.1
2	28.6	46.3	41.6	27.3
3	12.9	30.4	33.3	25.5
4	27.1	21.9	33.3	12.9

With children of different ages (or grades) combined, differences between the sexes for the four questions were not large. Dividing the group into first- and second-grade children versus third-graders, however, revealed significant differences. Table 2 shows the percentage of the children answering "Mother" to the four questions in relation to their grade-placement. For all four questions the older children were consistently more likely than the younger ones to see the same-sex parent as less benevolent and more frustrating. On Question 2, the third-grade boys viewed the mother as less punitive than the younger boys ($p = .07$). Similarly, on Question 3, significantly more third-grade girls saw the mother as boss than first- and second-grade girls ($p < .05$). One might assume that the parent who is perceived as "boss" is the major controller of gratifications and punishments and, therefore, the one who would be feared more. On Question 4, there was no large difference between the sexes for the younger children, but fear of the mother was significantly more frequent for the older girls than for the older boys ($p < .05$), supporting the trend toward perception of the same-sex parent as more threatening as the child grows older.

This trend toward a more threatening perception of the same-sex parent may have realistic bases and result, in part, from differential handling of boys and girls as they start school and begin to assume more definitive sex roles. In time the child may begin to perceive such a difference in parental attitude and treatment. A recent analysis (2) of interview data from mothers of preschool children showed that the mothers of girls were less nurturant at bed time and more frustrating at meals than mothers of boys of the same age, suggesting that a parent is less gratifying with a child of the same sex than with one of the opposite sex.

SUMMARY

Interviews were held with 217 children of both sexes whose ages ranged from 6–0 to 10–2. Their answers to questions related to their perception of their parents were obtained. Both girls and boys stated that fathers were less friendly and more dominant, punitive, and threatening than mothers. There was, however, a consistent tendency for the older children to be more likely than the younger children to view the parent of the same sex as more dominant and punitive. It was suggested that differential handling of boys and girls might partially account for this latter finding.

REFERENCES

1. FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. E. Perception of parents and social attitudes. Paper read at meeting of Amer. Psychopath. Soc., June, 1955.
2. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, No. 47.

Received September 13, 1955.

A FURTHER STUDY OF THE McKINNEY REPORTING TEST SCORES[1]

JOHN R. BARRY

Western Psychiatric Institute, University of Pittsburgh

RECENTLY, Douglass and Brown (4) reported on their investigation of reactions to frustration on the McKinney Reporting Test and concluded that the test measures of frustration reactions should be refined further. A study of these measures, utilizing a similar research design, was consequently undertaken as part of a large personality screening research program (7). The methodology and findings are described below and contrasted with those of Douglass and Brown.

The McKinney Reporting Test (5, 6) is a paper-and-pencil group test involving simple repetitive perceptual-motor responses to 315 items. The test consists of four comparable parts: Part I (the pre-frustration condition), Parts II and III (the frustration conditions), and Part IV (the post-

[1] This research was completed at the Department of Clinical Psychology, USAF School of Aviation Medicine, Randolph Field, Texas. Appreciation for their help is due to Drs. S. B. Sells and D. K. Trites, USAF School of Aviation Medicine, and to Dr. Roy M. Hamlin, Western Psychiatric Institute.

TABLE 1

Significance of Differences (p-Values*) Between Correlations of Part Scores in Present Study and in the Douglass-Brown (D-B) Study

Comparison of parts	Items attempted		Items correct		Percentage of items correct	
	Present study	D-B study	Present study	D-B study	Present study	D-B study
I–II	.07*	—	—	—	—	—
I–III	.03	—	—	—	.001*	.05
I–IV	.01	—	.01	—	—	—
II–III	.01	—	—	—	—	—
II–IV	.05	.05	.001*	—	—	—
III–IV	.001*	.01	.08	.01	—	—

* All p-values in this table are based upon a two-tailed test of significance. The asterisks indicate that the difference is inconsistent with the hypothesis that stress-sensitive subjects respond less consistently from part to part than stress-resistant subjects.

frustration condition). Three scores are obtained for each of the four parts: the number of responses made (correct plus incorrect), the number of responses correct, and the percentage correct (i.e., the number correct divided by the number made).

Douglass and Brown administered the test under the standard (frustrating) conditions to a group of 101 college students, and under nonfrustrating conditions to a similar group. It was hypothesized that the disorganizing effects of the frustrating conditions would be reflected in the lower intraindividual correlations among the part scores of the frustrated group when compared with the analogous correlations in the nonfrustrated group. Fifteen comparisons were made, utilizing the three scores described above. Four comparisons yielded statistically significant differences ($p < .05$) consistent with the research hypothesis. The authors concluded that the evidence for the sensitivity of the test to the frustrating conditions was not conclusive.

Procedure

In the present study, the McKinney Test was administered to 371 first pilots (Aircraft Commanders) and 428 copilots prior to their entering B-29 Combat Crew Training. Criterion ratings of each pilot's adjustment in training were developed to reflect anxiety-proneness, a predisposition to unfavorable reactions to stress and frustration, and generally poor defenses. The ratings were based upon medical and training records, sociometric data, and an eight-hour psychological evaluation, including an interview and psychodiagnostic testing. The training records included instructor's grades and comments, board proceedings, and the reasons why a student repeated or failed any parts of the program. These criteria were used by psychologists to make over-all ratings of adjustment on a seven-point scale. Intercorrelations among four raters varied from .85 to .95, indicating a high degree of rater agreement. These criteria are more completely described elsewhere (1, 2, 3).

A sample of 400 pilots, 200 of whom were upper and 200 lower criterion group members, was selected from the population of 799 pilots in such a way as to define extreme criterion groups. Specifically, the 100 first pilots who were rated as having made the best adjustment to training and the 100 rated as having made the poorest adjustment were designated as members of the upper and lower criterion groups respectively. In the same manner, 200 copilots were selected for the upper and lower groups. Thus, the two criterion groups were selected to define the maximum adjustment differences in the population.

It was predicted that the upper criterion group members would perform on the McKinney Test at a more consistent rate and would be less affected by the frustrating conditions of the test than the lower criterion group[2]. The lower criterion group members were expected to react in an unsystematic fashion during the stress parts of the test. Some of the lower group were expected to perform more efficiently and others more poorly under stress than the upper group. Thus, correlations among the part scores of the upper group were expected to be higher than those occurring in the lower group. Product-moment correlations between the part scores of upper and lower groups were compared by computing the critical ratios between z transformations of the correlations.

Results

Table 1 lists p values indicating the significance of the differences between correlations of the part scores. In the present study, the differences between the criterion groups were in the predicted direction and statistically significant ($p < .05$) for five of the 18 possible comparisons. For three additional comparisons, very significant differences ($p < .01$) in a direction *opposite* from that predicted were found. For two of these three comparisons, the Douglass-Brown data yielded statistically significant differences in a direction *opposite* from that found in the present study. In contrast to the findings in the present study, none of the statistically significant differences found by Douglass and Brown were inconsistent with their predictions. Thus, there appeared to be greater variability in the data of the present study, probably due in part to the criterion differences between the two groups and in part to the differences in the types of setting in which testing was done. In the present study, a partially successful effort was made to create the contrary-to-fact impression that each subject's test responses would have an influence on his subsequent military career. Thus, the test-taking attitudes of the military subjects

[2] No estimates of stress-sensitivity (unfavorable reaction to frustration) which were independent of the adjustment criterion were available. However, an inspection of the bases of the criterion ratings (3, pp. 4–8) suggested that real differences in frustration reactions should be found between the criterion groups.

probably were different from those of the college students in the Douglass-Brown study.

While it seems likely that the inconsistencies between the two studies may be partly due to differences in the research designs, it is also quite possible that the measures of frustration reaction reflected by both sets of data are unstable. Although certain aspects of the test appear to have frustrating effects,[3] various unidentified factors may be masking or interfering with the predicted relationships. The need for refinement in scoring and construction of the test, suggested by Douglass and Brown, is emphasized by these inconsistencies. Since in both studies there are significant differences in the predicted direction between the average intercorrelations of the experimental and control groups, both studies seem to support the experimental hypothesis that the test does enable real discrimination by means of frustration reactions between stress-sensitive and stress-resistant groups. The stress-sensitive persons as a group were less consistent in their test performance than the stress-resistant group.

SUMMARY

It was expected the stress-sensitive group would respond less consistently from part to part on this test than the stress-resistant group. The degree to which this was found is indicated in Table 1. Inconsistencies between the present study and an earlier one probably reflect the instability of the measures of frustration reaction as well as differences in the research designs. There appears to be a need for refinement in the scoring and construction of this test.

[3] In research reported elsewhere (1, 2), upper and lower criterion groups, similar to those described above, could be identified accurately by their responses to this test. Specifically, the upper group performed more efficiently than the lower group on all four parts of the test. The examiner-induced frustration in Parts II and III did not appear to sharpen the discrimination between the groups.

REFERENCES

1. BARRY, J. R., FULKERSON, S. C., & SELLS, S. B. Adaptability screening of flying personnel: Research on the McKinney Reporting Test, *USAF Sch. Aviat. Med. Proj. Rep.* (Rep. No. 56-5) (in press).
2. BARRY, J. R., FULKERSON, S. C., & SELLS, S. B. Prediction of adjustment with an "objective group frustration" test. *Amer. Psychologist*, 1955, **10**, 448–449. (Abstract)
3. BITTERMAN, M. E., HOLTZMAN, W. H., & BARRY, J. R. Psychiatric screening of flying personnel. Conditioning and extinction of galvanic skin response as a function of adjustment to combat crew training. *USAF Sch. Aviat. Med. Proj. Rep.* Feb., 1955 (Rep. No. 55-29).
4. DOUGLASS, W. B., & BROWN, K. B. Adequacy of the Reporting Test measures for detecting reactions to frustration. *J. abnorm. soc. Psychol.*, 1954, **49**, 593–595.
5. FULKERSON, S. C. Individual differences in reaction to failure-induced stress. *Amer. Psychologist*, 1955, **10**, 336–337 (Abstract)
6. MCKINNEY, F., STROTHER, G. B., HINES, RUTH R., & ALLEE, RUTH A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.
7. SELLS, S. B., & BARRY, J. R. A research program to develop psychiatric selection of flying personnel. I. Theoretical approach and research design. II. Research progress. *J. aviat. Med.*, 1953, **24**, 29–47.

Received October 13, 1955.

SOCIAL ORDER

"... is stimulating and forward looking."
—*Family Life Institute*

"... intensely interesting."
—*Dorothy Day*

Current and Coming Titles

FRANCIS WILSON:
Intellectualism and Religion

JOSEPH H. FICHTER:
Background of Religious Functionaries

WILLIAM FERREE:
The Role of Social Justice

LEO C. BROWN:
Unionism in Small Plants
(1955 Hillman Lecture)

JOHN L. THOMAS
The Social Web of Marriage

MAURUS BARRANECHEA:
The Labor Schools and the Future

ERNEST W. BURGESS:
American Catholic Family

ARTICLES · TRENDS · REVIEWS
one year $4 — two years $7 — three years $9

SOCIAL ORDER
Desk A – A-13
3908 Westminster Place St. Louis 8, Mo.

INTERPERSONAL DIAGNOSIS OF PERSONALITY*

- MULTILEVEL CLINICAL DIAGNOSIS AND PROGNOSIS
- ANALYSIS OF GROUP DYNAMICS IN PSYCHOTHERAPY OR INDUSTRIAL MANAGEMENT
- FAMILY DIAGNOSIS IN CHILD GUIDANCE AND MARITAL COUNSELING
- MEASUREMENT OF CHANGE IN PSYCHOTHERAPY
- PERSONALITY RESEARCH IN ANY OF THESE AREAS

TEST MATERIALS NOW AVAILABLE

INTERPERSONAL CHECK LIST. A 128-item test by means of which the subject diagnoses himself, family members, fellow group members, therapist, etc. Also used by others to rate the subject.

RECORD BOOKLET FOR INTERPERSONAL DIAGNOSIS OF PERSONALITY. Clinical diagnosis and prediction of interpersonal behavior.

RECORD BOOKLET FOR ANALYSIS OF GROUP DYNAMICS. Interpersonal perceptions by each member in a group of every other member. Diagnosis of group dynamics—overt and covert patterns of relationships.

RECORD BOOKLET FOR FAMILY DIAGNOSIS. Multilevel diagnosis of husband and wife, their misperceptions of each other and of their children. Oedipal structure of the marriage.

MANUAL FOR INTERPERSONAL DIAGNOSIS

The manual provides a step-by-step description of all the procedures for interpersonal diagnosis. Administration and scoring of each interpersonal test is outlined. All steps involved in standardization, plotting of scores, interpersonal diagnosis and the calculation of conflict indices are discussed and illustrated.

The use of the interpersonal system in research is given detailed consideration. Particular attention is given to the handling of multilevel data in personality research.

*Developed by Leary et al. at the Kaiser Foundation, Oakland, California. The Interpersonal System has been used in its prepublication form by over 60 institutions. See:

Interpersonal Diagnosis of Personality, Ronald Press, 1956
J. Pers., 1951, *20*, p. 144; 1954, *23*, p. 129; 1955, *24*, p. 96; 1956, Dec.
Op. Psychother., 1954, *7*, p. 7.
J. Abnorm. Soc. Psychol., 1955, *50*, p. 110.
Psychiat., 1955, *18*, p. 147.
J. Clin. Psychol., 1956, *12*, p. 122.

A descriptive catalog and sample copies of these materials can be obtained from:

PSYCHOLOGICAL CONSULTATION SERVICE
1230 Queens Road
Berkeley 8, California

WILEY BOOKS

THE PSYCHOLOGY OF OCCUPATIONS*

By ANNE ROE, Research Psychologist. Builds a picture of the relations between occupations and other aspects of life, searching for a general pattern and for underlying principles. This remarkable new book provides a clear compendium of the known data on the differences in people that suit or disqualify them for given occupations.

Based on an extended study of the relationships that exist between vocations and personality, the book is not confined to any single field or aspect. An entirely new classification of occupations is used as a foundation for this study. Its findings are, at the same time, general and precise. The work is sufficiently broad to include discussions of such important factors as social background and the all-important relationship between occupational choice and satisfaction. *1956. 340 pages. $6.75.*

THE ANALYSIS OF FANTASY*

The Thematic Apperception Technique in the Study of Personality

By WILLIAM E. HENRY, University of Chicago. Gives a comprehensive statement of the theory of projective interpretation with a practical, down-to-earth discussion of the categories and methods of analysis, backed up by ample illustrations drawing on normal case material. Outlined are the major areas of personality upon which Thematic Apperception Techniques can provide data. Also included is a systematic plan for relating these conceptual areas to variables in the TAT response.

The book permits you to build a coherent system for procedure in interpretation by giving detailed descriptions of the variable of response (form, content, negative content, etc.). It emphasizes the importance of social factors and social setting in the study of personality dynamics. Its extensive bibliography, currently the most comprehensive one available, is arranged for maximum convenience by topics of interest. *1956. 305 pages. $6.00.*

OPINIONS AND PERSONALITY

By M. BREWSTER SMITH, Social Science Research Council; and JEROME S. BRUNER and ROBERT W. WHITE, both of Harvard University. Looks at the lives and characters of ten normal men in the midst of their adult lives and asks: What function is served in each life by holding a particular set of opinions about Russia and Communism? By re-examining various classic problems of opinion research, the work helps to clarify the very nature of opinion and presents opinion in the context of lives in process. The results of these extensive case studies serve to reconcile the one-sided accounts often given by psychoanalysis, sociology, trait psychology, and cognitive theory, bringing them into balanced perspective. These portraits also emphasize the importance of constructive processes in the functioning of personality, and the key role played by attitudes and values in these processes. *1956. 294 pages. $6.00.*

★ *One of the Wiley Publications in the MENTAL HEALTH SCIENCES.*

Send for examination copies.

JOHN WILEY & SONS, Inc. 440 4th Ave., New York 16, N.Y.